VOLUME 102 PART B NUMBER 3 MAY 1955

The Proceedings OF THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871: INCORPORATED BY ROYAL CHARTER 1921

PART B

RADIO AND ELECTRONIC ENGINEERING
(INCLUDING COMMUNICATION ENGINEERING)

SAVOY PLACE . LONDON W.C.2

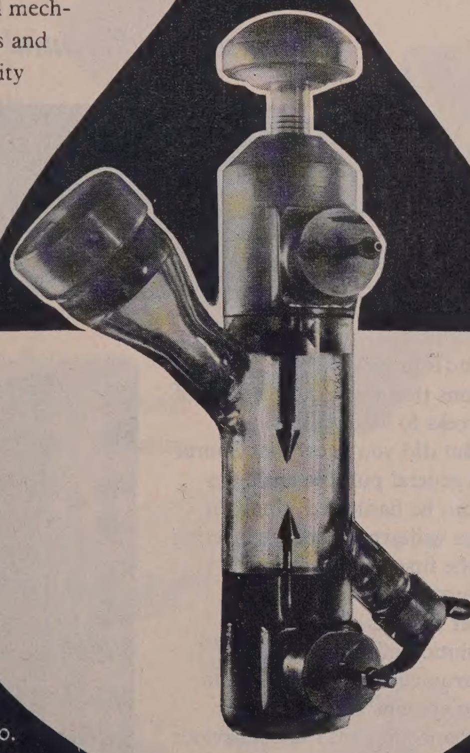
Price Seven Shillings and Sixpence

Cast in ARALDITE

This component is a part of the 'Agglometron'* oil treatment chamber.
It was cast in 'Araldite' because:—

*An outstanding electrical insulator was needed
*It had also to be oil-resistant
*It was important that it should lend itself readily to casting and machining

'Araldite' as a casting and as a bonding resin is used in many components of the 'Agglometron'. Its electrical and mechanical properties, its exceptional adhesion to metals and ceramics, its resistance to high temperatures, humidity and corrosive agents suggest other uses in which the execution of new designs can be made practicable and the production of electrical components greatly simplified. May we send you full descriptive literature?



* An electrical device developed by Mr. O. E. Nekolla of Messrs. Menrow Ltd. (a subsidiary company of J. & E. Arnfield Ltd.) for agglomerating impurity particles in circulating oil to facilitate filtering. The name 'Agglometron' is a registered trade mark for the high voltage treatment chamber included in the filtration plant and is covered by B.P. 698,900.

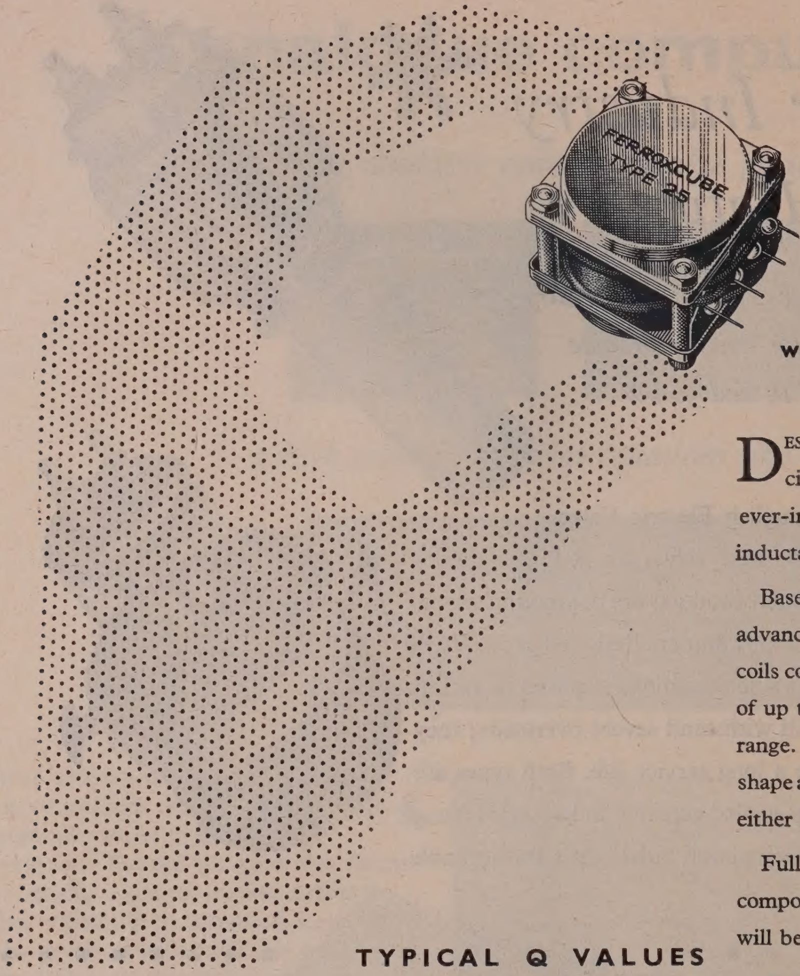
THESE ARE THE NEW EPOXIES!

'Araldite' (regd.) epoxy resins are obtainable in the following forms:

- *Hot and cold setting adhesives for metals and most other materials in common use.*
- *Casting Resins for the electrical, mechanical and chemical engineering industries.*
- *Surface Coating Resins for the paint industry and for the protection of metal surfaces.*

'Araldite' *epoxy casting resins*

Aero Research Limited *DUXFORD, CAMBRIDGE. Telephone: Sawston 187*
A Ciba Company



High Q inductance coils

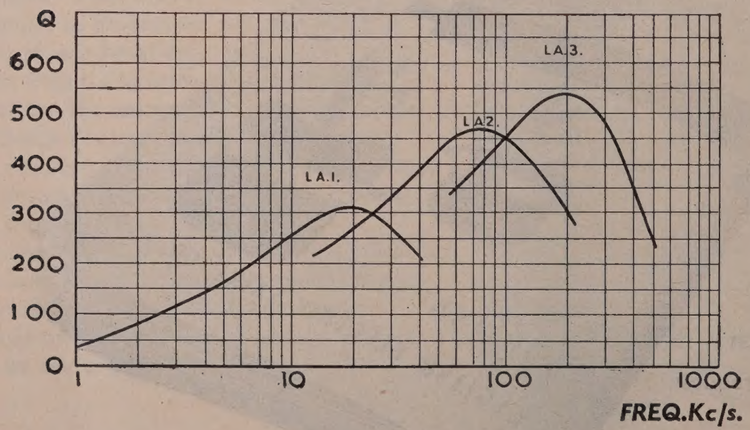
wound on Ferroxcube cores

DESIGNERS of compact and efficient tuned circuits and wave filters are making ever-increasing use of Mullard high Q inductance coils.

Based on Ferroxcube, the world's most advanced magnetic core material, these coils combine small size with an inductance of up to 30 henries over a wide frequency range. Furthermore, their convenient shape and self screening properties facilitate either individual mounting or stacking.

Full details of these and other high grade components now available from Mullard will be gladly supplied on request.

TYPICAL Q VALUES



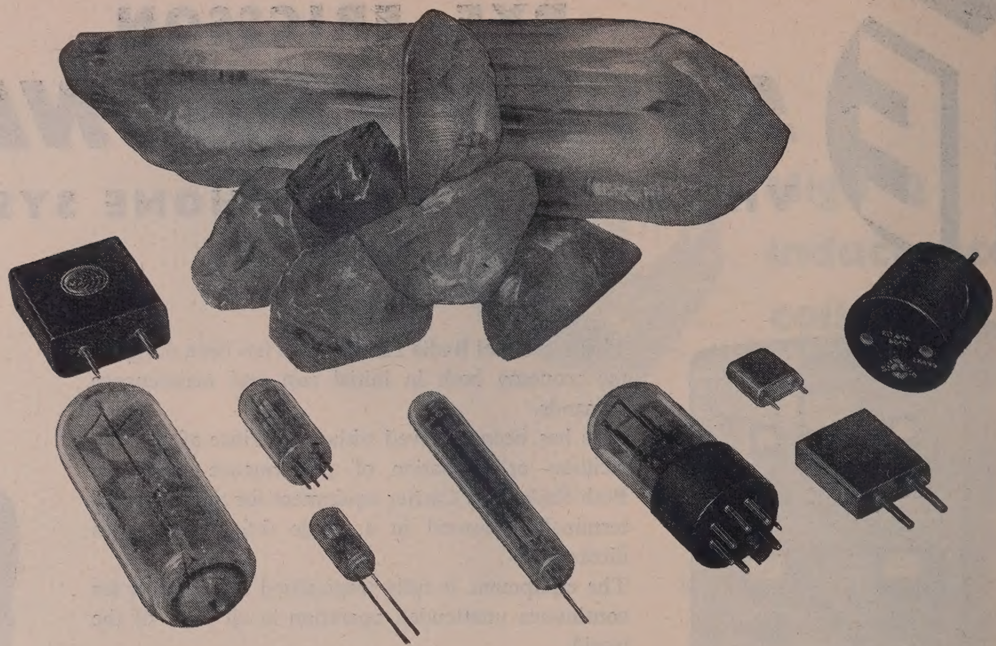
Special Features

Small size
Low hysteresis loss factor
High value of inductance
Low self capacitance
Controllable air gap facilitating inductance adjustment
Self screening
Controlled temperature coefficient
Operation over a wide frequency range
Easily mounted

Mullard

'Ticonal' permanent magnets,
'Magnadur' ceramic magnets,
Ferroxcube magnetic cores.

MULLARD LIMITED · COMPONENT DIVISION · CENTURY HOUSE · SHAFTESBURY AVENUE · LONDON · W.C.2
(MM450)



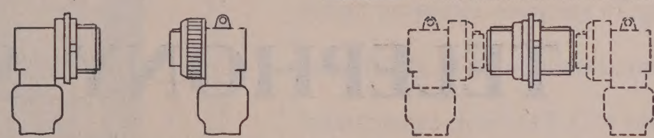
For special purpose connectors

...contact Plessey

Plessey have long specialised in the design and mass production of miniature electrical connectors of high reliability.

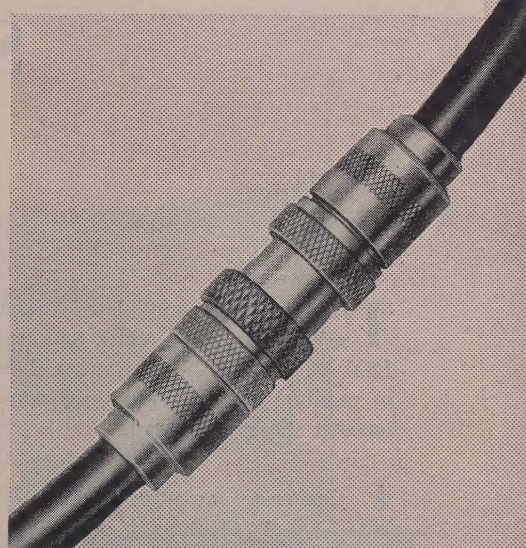
In particular, Extra High Tension connectors having high insulation properties and suitable for high voltages, are available. There are two types, Demountable (7kV. peak) as illustrated and Moulded (10 kV. peak) as shown in outline. Both are interchangeable and units of each are obtainable for free cable, bulkhead or panel installations.

EHT CONCENTRIC TYPES



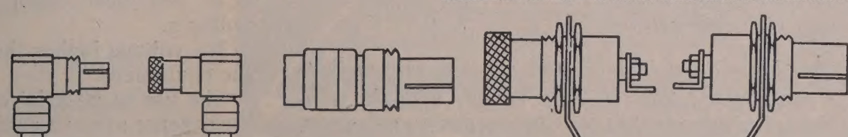
Panel plug or socket Cable plug or socket Bulkhead connectors

SHF TYPES



Plessey also manufacture a range of highly efficient interchangeable connectors for the effective and reliable termination of Uniradio Cable operating equipment in the Super High Frequency bands (above a hundred megacycles and up to 10,000 megacycles).

Basically designed for 70-80 ohm lines these units are standardised so that no confusion can arise when making connections between various S.H.F. devices.



Cable plug minor Cable socket minor Cable plug major Panel socket major Panel plug major

These Publications are available to manufacturers on request . . .

Publication No. 560/1 on Plessey E.H.T. concentric connections.
Publication No. 672/1 on Plessey S.H.F. connections

THE PLESSEY COMPANY LIMITED · ILFORD · ESSEX

PB2b

JAMES NEILL & CO. (SHEFFIELD) LTD.
SHEFFIELD 11 ENGLAND

★ Made by the makers of "Eclipse" Permanent Magnet Chucks

M2

In Science and Industry alike—among technicians, manufacturers and those engaged in the sale of electrical products—as well as among the public at large, the Philips emblem is accepted throughout the World as a symbol of quality and dependability.

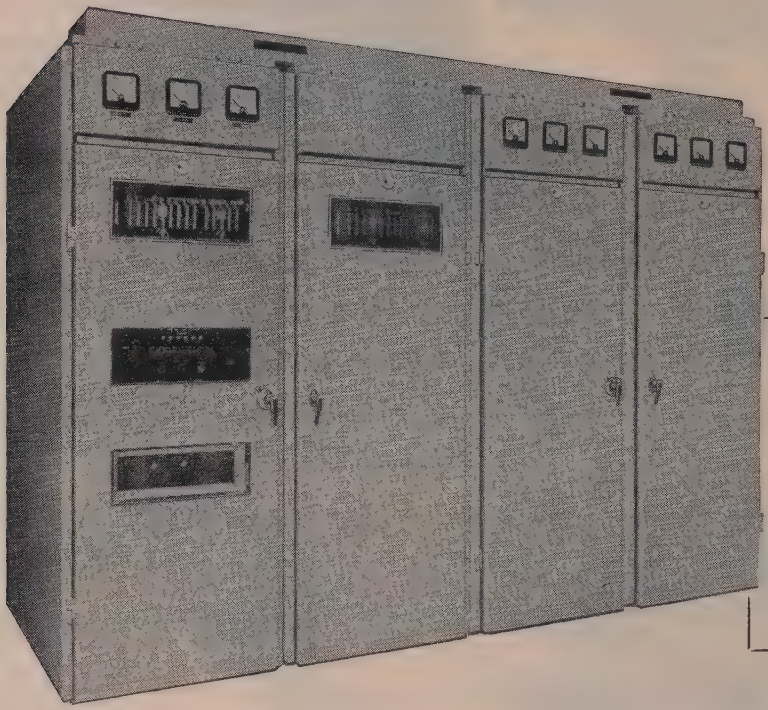
RADIO & TELEVISION RECEIVERS · TUNGSTEN, FLUORESCENT, BLENDED & DISCHARGE LAMPS & LIGHTING EQUIPMENT · 'PHILISHAVE' ELECTRIC DRY SHAVERS · 'PHOTOFLUX' FLASHBULBS · HIGH-FREQUENCY HEATING GENERATORS · X-RAY EQUIPMENT FOR ALL PURPOSES ELECTRO-MEDICAL APPARATUS · ARC & RESISTANCE WELDING PLANT AND ELECTRODES · ELECTRONIC MEASURING INSTRUMENTS · MAGNETIC FILTERS · BATTERY CHARGERS & RECTIFIERS · SOUND AMPLIFYING INSTALLATIONS · CINEMA PROJECTORS · RECORDING APPARATUS

PHILIPS ELECTRICAL LTD.

CENTURY HOUSE, SHAFTESBURY AVENUE, LONDON, W.C.2

P.143A

Marconi 6kW HF ISB Transmitters



TYPES
HS 71 AND HS 72

The assembly is enclosed by unit sections, as shown here, with access through front and rear doors. The two left hand bays house the rectifier and power equipment and the right hand bays the low power and auxiliary transmitting circuits and the main output stage.

These transmitters, designed in accordance with the most advanced practice, provide :—

(a) Telegraphy on CW and FSK (A1 and F1)
(b) Independent Sideband Operation (A3b)

The drive equipment is external and provides either ISB modulation or telegraph keying at 3.1 Mc/s and suitable RF oscillator signals for frequency changing in the transmitter. HS 71 is manually operated; HS 72 provides full automatic tuning and selection of any one of six pre-set frequencies.

FEATURES INCLUDE

- **Tuning over the whole range without change of components**
- **Air cooling throughout, with dust filtering.**
- **Double screening of power stages reduces indirect radiation and cooling air noise.**
- **Envelope feed back to reduce distortion.**
- **Compact assembly with good access for servicing and safety interlocking.**

More than 80 countries now have Marconi equipped telegraph and communication services, many of which, completed 20 years ago, still give trouble-free operation.

Lifeline of communication

MARCONI

COMPLETE COMMUNICATION SYSTEMS

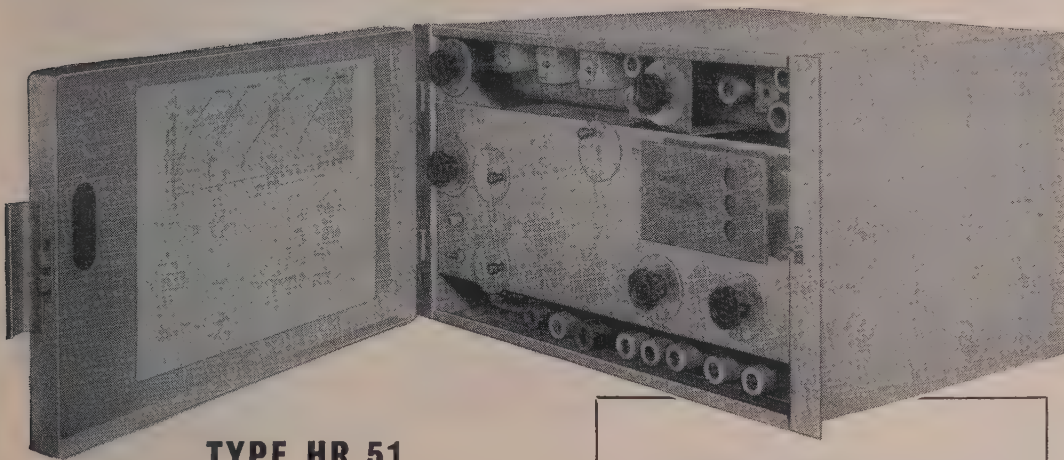
Surveyed, planned, installed, maintained

Partners in progress with The 'ENGLISH ELECTRIC' *Company Limited*

MARCONI'S WIRELESS TELEGRAPH CO. LTD., CHELMSFORD, ESSEX

LC12

Marconi Three Channel Telegraph/Telephone H.F. Receiver for remote control



TYPE HR 51

The Type HR 51 equipment is suitable for reception of telegraph or telephone signals on any one of three pre-set HF channels. It may be remotely controlled for channel selection and fine tuning from a distance of up to 10 miles. Control can be over the same wires as carry the AF signal output or a separate pair, provided the control circuit does not exceed 1000 ohms loop resistance. The receiver can be used to operate a recording unit such as the Marconi HU 11. Two may be connected for diversity reception, feeding a recording unit such as the Marconi HU 12.

Power supply components are housed in a compact bench mounting cabinet with the receiver. Access to all receiver controls and the valves is by a hinged door, which protects the controls from accidental interference. Lamps indicating the selected channel are visible through an aperture when the door is shut. Further access is by removable panels. The HR 51 is also available for rack mounting with associated equipment. The remote control unit, not shown here, is suitable for bench or rack mounting and requires connection to a mains supply point.

Over 80 countries now have Marconi equipped communications systems. Many of these are still giving trouble free service after more than twenty years in operation.

Lifeline of communication

MARCONI

COMPLETE COMMUNICATION SYSTEMS

Surveyed, planned, installed, maintained

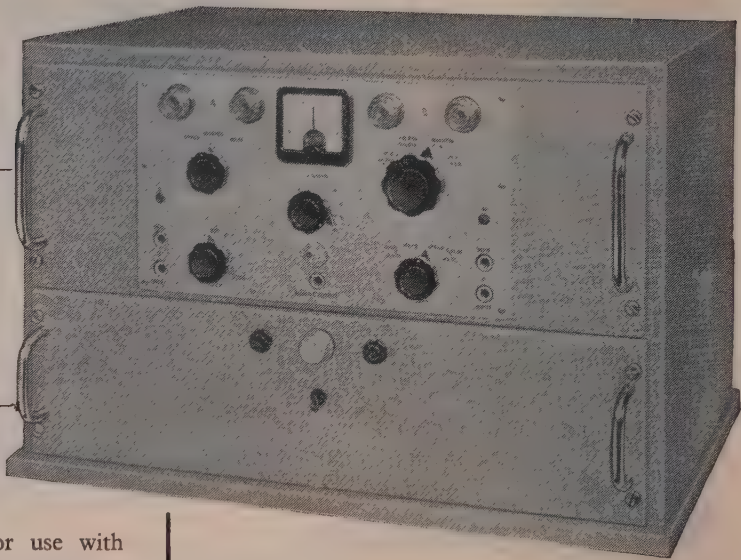
Partners in progress with The 'ENGLISH ELECTRIC' *Company Limited*

MARCONI'S WIRELESS TELEGRAPH CO. LTD., CHELMSFORD, ESSEX

LC 14

Marconi Double Diversity Telegraph Recording Unit

TYPE HU.12



Shown here is the HU.12, in the bench mounting cabinet where it fits on top of the power unit. It can also be built into a standard rack which houses the associated receivers.

The HU.12 unit is designed for use with two suitable receivers of frequency shift or on-off telegraph transmissions in double diversity. The combined AF outputs of the receivers are converted into double current signals to operate an undulator, relay or other 'space-mark' type of equipment requiring a current up to 30 mA. No receiver modification is necessary except, perhaps, adjustment of the BFO frequency.

Satisfactory recording is possible should the first oscillator drift up to ± 800 c/s.

FEATURES :-

- **Electronic path diversity selector for FSK will function when difference between paths exceeds 4 db.**
- **Alternative DC output suitable for a low resistance undulator (20+20 ohms) is provided.**
- **Neon indicators for visual tuning of FSK reception.**
- **Front of panel control for pre-set 'mark' and 'space' current adjustment, filter selection and signal bias.**

Over 80 countries now have Marconi equipped telegraph and communication systems. Many of these are still giving trouble-free service after more than 20 years in operation.

Lifeline of communication

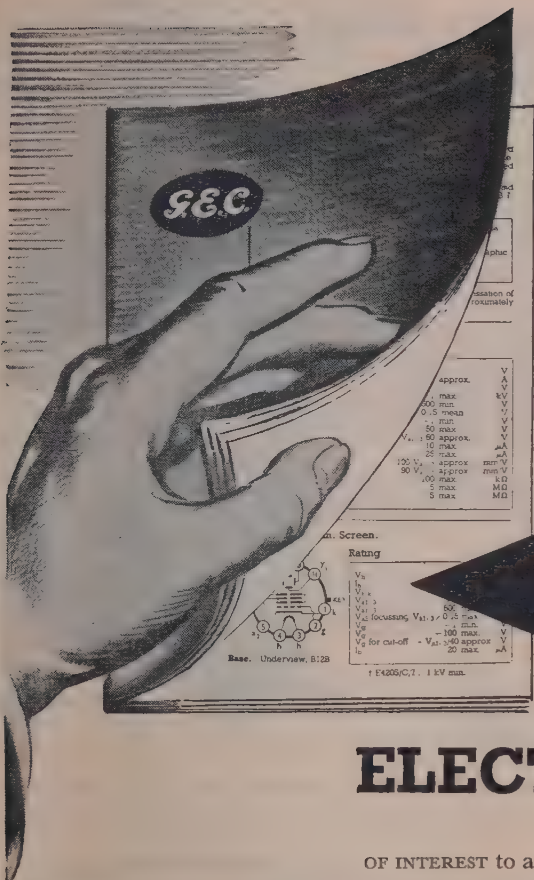
MARCONI

COMPLETE COMMUNICATION SYSTEMS

Surveyed, planned, installed, maintained

MARCONI'S WIRELESS TELEGRAPH CO. LTD., CHELMSFORD, ESSEX

Partners in Progress with The **'ENGLISH ELECTRIC'** *Company Limited* LC 13



ELECTRONIC DEVICES

OF INTEREST to all engineers, this new publication forms an invaluable reference to current G.E.C. electronic products suitable for such industrial applications as control instruments, scientific measurement, counting, radio-active indication, photoelectric alarms and indicating devices and electronic relays and voltage stabilisation. The technical information which includes ratings, dimensions, base connection diagrams and screen characteristics, enables the potential user to choose the device most suitable for a particular application.

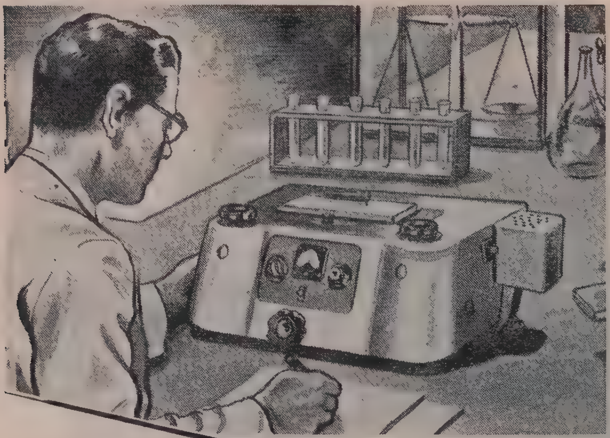
The contents include:

- Barretters
- Cathode ray tubes (instrument and special types)
- Electrometer valves
- Geiger Muller tubes
- Photoelectric cells
- Photoelectric industrial aids
- Semi-conductors
- Voltage stabilisers

Write to

OSRAM VALVE AND ELECTRONICS DEPT. FOR **OV 2507**

THE GENERAL ELECTRIC CO. LTD., MAGNET HOUSE, KINGSWAY, LONDON, W.C.2



the Research Chemist relies upon

CONSTANT VOLTAGE....

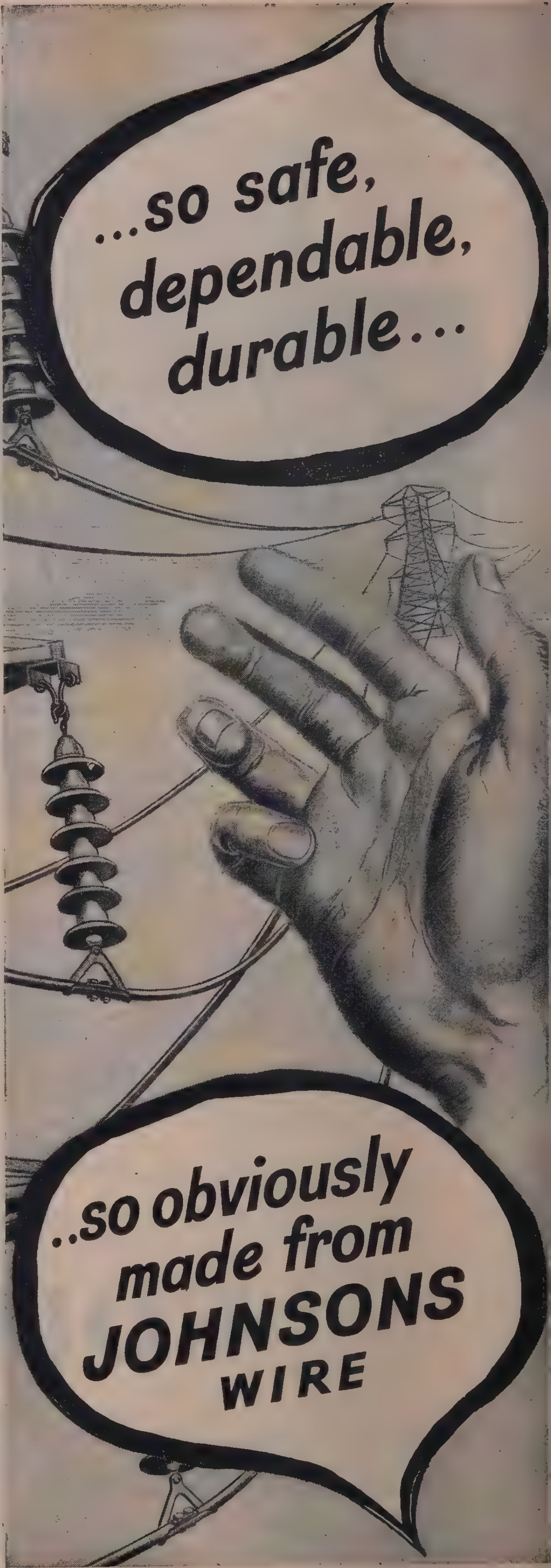
In problems of analysis to-day, the chemist is aided to an increasing degree by electronics. However, spectroscopic techniques and apparatus depend for their accuracy on a reliable constant voltage supply, and here Advance Constant Voltage Transformers provide the answer. Purely automatic, and with no thermionic valves or moving parts, they can be relied upon to provide the stabilized A.C. voltage so necessary in the laboratory.

Full details given in folder R28 gladly sent on request.

(Illustration of spectrophotometer by courtesy of Unicam Instruments.)

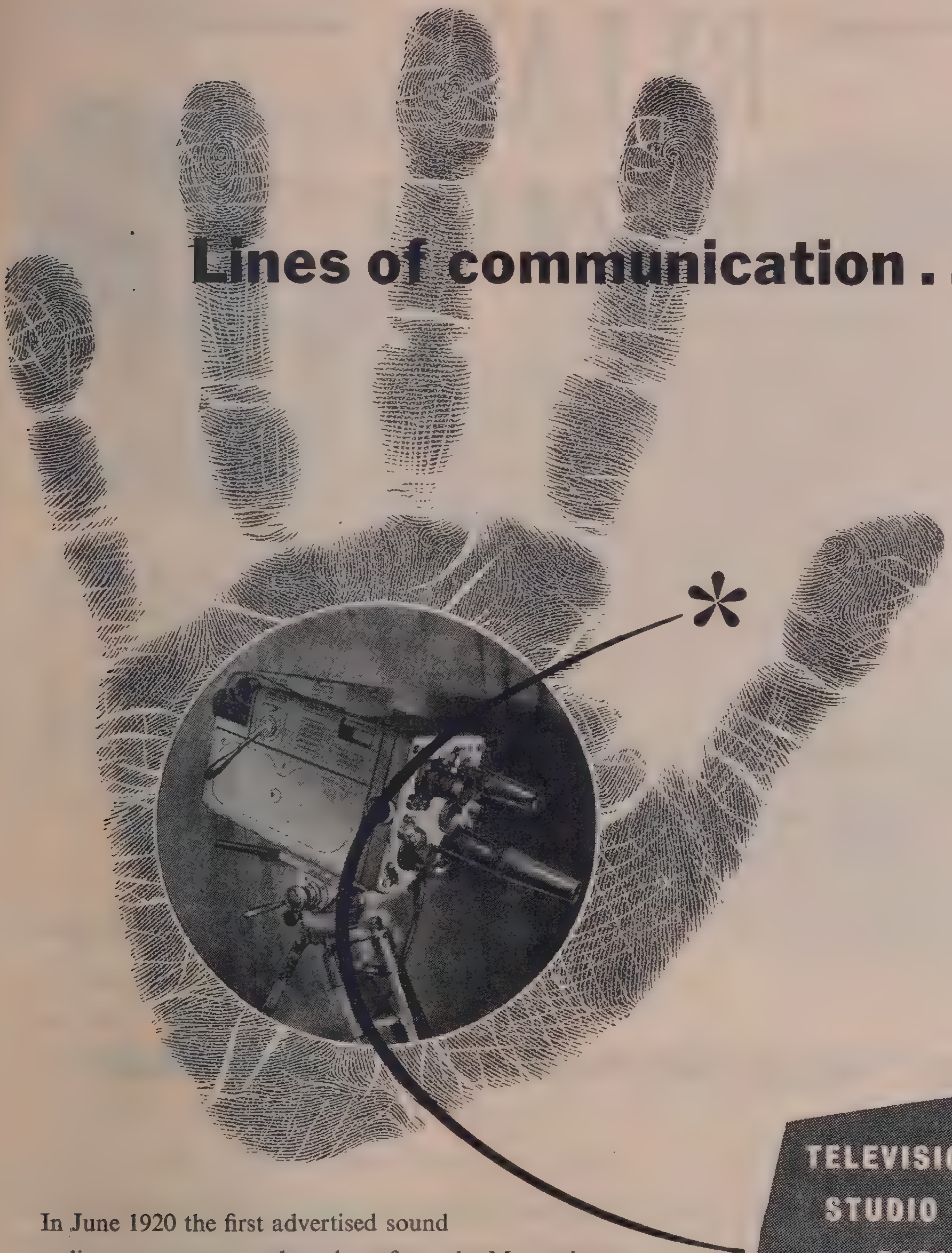
Advance
CONSTANT VOLTAGE TRANSFORMERS

CVT ADVANCE COMPONENTS LTD.
MARLOWE ROAD, WALTHAMSTOW, LONDON, E.17
Telephone: LARkswood 4366/7/8



RICHARD JOHNSON & NEPHEW LTD.
FORGE LANE MANCHESTER 11

Lines of communication . . .



In June 1920 the first advertised sound radio programme was broadcast from the Marconi transmitter at Chelmsford. To-day Marconi high or medium power transmitters and high power aerials are installed in every one of the B.B.C's television transmitter stations and Marconi television systems are being supplied to countries in North and South America, Europe and Asia.

TELEVISION CAMERAS
STUDIO EQUIPMENT
RADIO LINKS
AND TRANSMITTERS
AERIAL SYSTEMS
COLOUR TELEVISION
INDUSTRIAL TELEVISION

MARCONI

Complete Television and Sound Broadcasting Systems

MARCONI'S WIRELESS TELEGRAPH COMPANY LIMITED · CHELMSFORD · ESSEX

LG 3

RELAYS

SERIES 2400

A Relay of noteworthy dimensions, designed in size and performance to suit present-day electronic equipment. The new 2400 Relay is available with twin light duty or single heavy duty contacts. When fitted with a 10,000 ohm coil, the pull-in is approximately 4 milliamperes; contact pressure and clearance have not been sacrificed to achieve this sensitivity.

DIMENSIONS: Above chassis $2\frac{1}{2}''$ high x $1''$ wide x $1\frac{5}{8}''$ deep.

WEIGHT: $4\frac{1}{2}$ ounces.

Telephone: Newmarket 3181-2-3

Telegrams: Magnetic, Newmarket

MAGNETIC DEVICES LIMITED

MAGNETIC DEVICES LTD
EXNING ROAD, NEWMARKET

MD 3A·2

ARE YOUR ENGINEERS OVERWORKED?

We have well-equipped laboratories, backed by production facilities, for the design, engineering and production of specialized electronic equipment.

A very wide range of microwave test equipment is available, including first grade test benches. Fully automatic test and calibration equipment, frequency and power sub-standards, technical investigation of Service equipment, are among the projects handled.

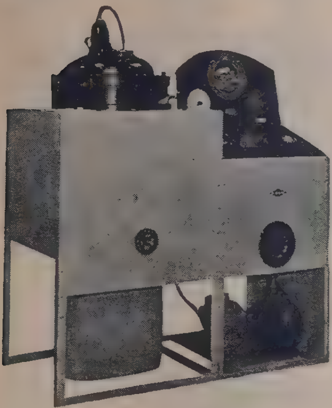
Why not discuss your problems with us?

G. & E. BRADLEY LIMITED
ELECTRONIC RESEARCH ENGINEERS

BERESFORD HOUSE, MOUNT PLEASANT, ALPERTON, WEMBLEY, MIDDX. TELEPHONE: WEMBLEY 3691–4
A.I.D. Approved. Contractors to H.M. Government.

IMPREGNATE
your coils with ease
BLICKVAC

High Vacuum Impregnators meet the most stringent specifications and yet are easy to handle. Full range of models available to meet the needs of the large-scale producer, the research laboratory or the small Rewind shop.



Outstanding Features:

- Ease in control
- Ease in cleaning
- Elimination of vibration
- Unequalled flexibility and performance
- Simple attachment of auxiliary autoclaves
- Units available suitable for Varnish, Wax, Bitumen and Potting Resins. Users include M.O.S., N.C.B., G.E.C., Pye, Marconi, Metro-Vick.

If your problem is COIL IMPREGNATION or impregnating or casting with Potting Resins consult:

BLICKVAC ENGINEERING LTD

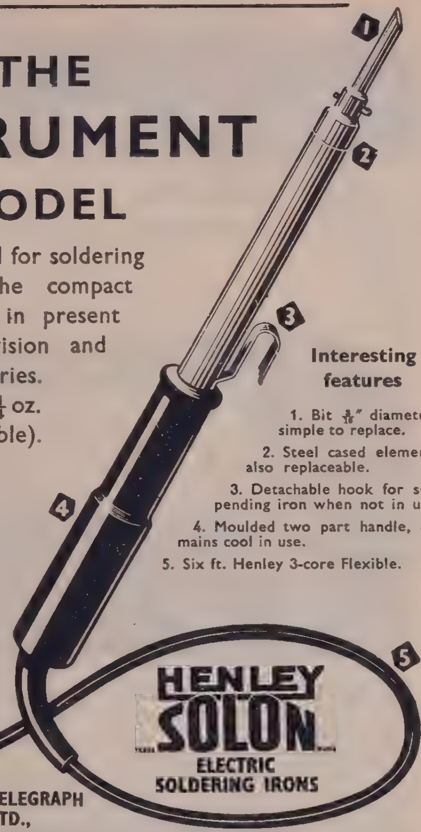
Bede Trading Estate, Jarrow, Co. Durham — Jarrow 89–7155
96–100 Aldersgate Street, London, E.C.1. — Monarch 6256/8

THE INSTRUMENT MODEL

Specially designed for soldering operations in the compact assemblies used in present day radio, television and electronic industries.

Weight $3\frac{1}{4}$ oz.
(excluding flexible).

Length 9 in.
25 Watts—
200/220 volts or
220/240 volts.



Interesting features

1. Bit $\frac{3}{16}$" diameter, simple to replace.
2. Steel cased element, also replaceable.
3. Detachable hook for suspending iron when not in use.
4. Moulded two part handle, remains cool in use.
5. Six ft. Henley 3-core Flexible.

HENLEY SOLON ELECTRIC SOLDERING IRONS

W. T. HENLEY'S TELEGRAPH WORKS CO., LTD.,
51/53, Hatton Garden, London, E.C.1

ZENITH
(REGD. TRADE-MARK)

PHASE SHIFTING TRANSFORMER

This instrument provides convenient means for adjusting the phase angle or power factor in alternating current circuits when testing single or polyphase service meters, wattmeters, or power factor indicators, etc. It is also the simplest means for teaching and demonstrating Alternating Current Theory as affecting phase angle and power factor.

Illustrated brochure free on request

The ZENITH ELECTRIC CO. Ltd.
ZENITH WORKS, VILLIERS ROAD, WILLESDEN GREEN
LONDON, N.W.2

Telephone: WILlesden 6581-5 *Telegrams:* Voltaohm, Norphone, London

MANUFACTURERS OF ELECTRICAL ENGINEERING PRODUCTS INCLUDING RADIO AND TELEVISION COMPONENTS

. . on R/F
INDUCTION HEATING

. . . a new publication of outstanding importance to all interested in brazing or heat treating.

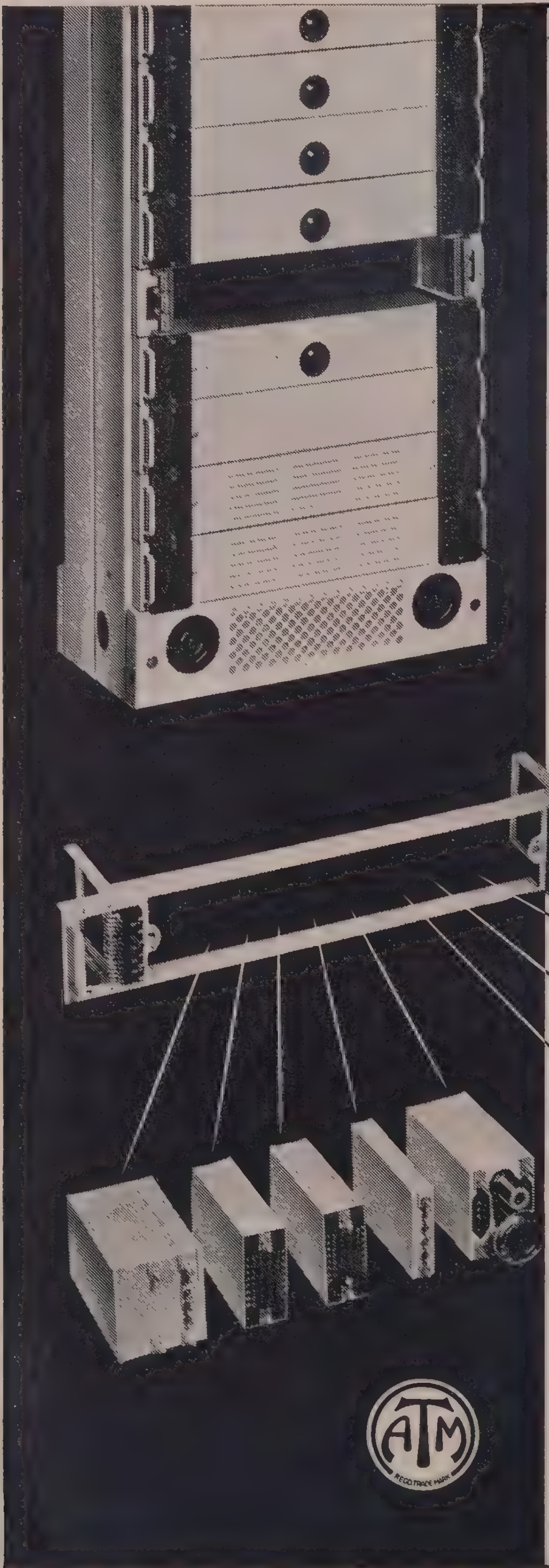
WRITE ON YOUR LETTER HEADING FOR YOUR FREE COPY OF PUBLICATION IH 501. ADDITIONAL COPIES 6/- EACH.

radio heaters ltd., wokingham, berks.


26
FERRANTI
220kV POWER TRANSFORMERS
for NEW ZEALAND
The State Hydro-Electric Department of New Zealand has ordered 26 Ferranti Large Power Transformers of the 220 kV class. 13—14,800 kVA, single-phase, 50 cycles, 11/220 kV Transformers forming 4—44,400 kVA 3-phase banks have already been supplied to ROXBURGH. 13—9,259 kVA, single-phase, 50 cycles, 11/220 kV Transformers forming 4—27,777 kVA 3-phase banks are under construction for WHAKAMARU.
WHAKAMARU
ROXBURGH
FERRANTI
FERRANTI LTD · HOLLINWOOD · LANCS
London Office: KERN HOUSE, 36, KINGSWAY W.C.2

FT145


CATHODEON
Quartz Crystals
FOR RELIABLE FREQUENCY CONTROL
Prompt Delivery, all types, 2,000–20,000 kc/s
When ordering 10X replacements, why not use our hermetically sealed Type 2XL?
CATHODEON CRYSTALS LIMITED
LINTON · CAMBRIDGESHIRE · Telephone LINTON 223



THE PROCEEDINGS OF
THE INSTITUTION OF ELECTRICAL ENGINEERS

EDITED UNDER THE SUPERINTENDENCE OF W. K. BRASHER, C.B.E., M.A., M.I.E.E., SECRETARY

VOL. 102. PART B. No. 3. MAY 1955

621.316.722.1.027.3 : 621.311.6 : 621.385.833

Paper No. 1729 M
Oct. 1954

THE USE OF AN ELECTRON VELOCITY ANALYSER TO STABILIZE A 50-kV DIRECT-VOLTAGE SOURCE TO A FEW PARTS IN A MILLION

By M. E. HAINE, M.Sc., F.Inst.P., Member, and M. W. JERVIS, M.Sc.Tech., Associate Member.

(*The paper was first received* 30*th March, and in final form* 1*st July*, 1954. *It was published in October*, 1954, *and was read before the* MEASUREMENTS SECTION, 14*th December*, 1954.)

SUMMARY

A new method of high-voltage stabilization and its practical application are described. The method uses an electron velocity analyser to measure changes in voltage. The operation and limitations of the device are analysed in detail.

The degenerative feedback circuit includes parallel d.c. and a.c. amplifiers. In operation, a stability of 2–3 parts in 10^6 is obtained on a 50-kV supply.

LIST OF SYMBOLS

S = Reciprocal of the fractional high-voltage stability.
N = Reciprocal of the resistance potential-divider ratio.
A_a = D.C. amplifier gain.
A_m = H.V. supply modulator gain.
V_i = Input voltage fluctuations, volts.
V_R = Reference potential, volts.
x_D = Diameter of the electron orbit, cm.
V = Electron accelerating potential, volts.
B = Flux density of magnetic deflecting field, G.
V' = Voltage corresponding to overlap distance, volts.
i = Fractional change in beam current.
g = Analyser sensitivity, amp/volt.
I' = Maximum value of collected overlap current, amp.
α = Semi-angle of electron beam, rad.
ω = Angle of marginal ray to XY plane, rad.
γ = Angle of marginal ray to YZ plane, rad.
I = Electron beam current, amp.
J_0 = Current density at centre of spot, amp/cm^2.
J_c = Current density at the cathode, amp/cm^2.
r_0 = Half-width of Gaussian distribution, cm.
β = Gun brightness, amp/cm^2/steradian.
T = Absolute temperature of the cathode, °K.
k = Boltzmann's constant (1/11 600 eV/°K).

(1) INTRODUCTION

For an experimental attempt to achieve a resolving power of atomic dimensions with the electron microscope,[2] a 50-kV d.c. power supply, stable to a few parts in 10^6 over time intervals up to at least ten minutes, was required. The paper describes considerations leading to the selection of a promising method of satisfying this requirement, and the development and operation of the apparatus.*

(2) CHOICE OF STABILIZING SYSTEM

The degenerative feedback method, commonly used to stabilize a high direct voltage, is shown diagrammatically in Fig. 1.

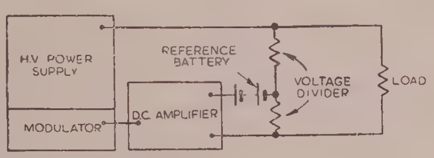


Fig. 1.—Block diagram of conventional voltage-stabilizing circuit.

A generator for the high voltage, e.g. a transformer-rectifier arrangement, feeds the load. A measure of the high voltage is obtained from a tap on a resistance potential-divider. The voltage from the tap to earth is fed to the input of the d.c. amplifier via a reference battery which backs-off the tap voltage to a suitable level for the amplifier input. The output of the amplifier modulates the high voltage in the sense necessary to correct any residual fluctuations present.

Stabilizers of this type have been described by previous authors[3,4,5] and, in general, stabilities to about 1 part in 10^4 are obtained.

There are three main requirements for a degenerative stabilizer which is to give an output stability within 1 in S:

(*a*) The total feedback gain given by the product of the potential-divider ratio ($1/N$), the amplifier gain (A_a) and the modulator gain (A_m) must exceed the ratio of the fractional fluctuations occurring before stabilization is applied ($\delta V_i/V_i$), including those due to load changes to the maximum permissible output fluctuations ($(\delta V/V = 1/S)$, i.e.:

$$A_a \times A_m/N > S\delta V_i/V_i$$

* A brief description of the principles of the stabilizer was given at the International Conference on Electron Microscopy held in Paris, September 1950. (Ref. 1.)

The authors are in the Research Laboratory of Associated Electrical Industries, Ltd.

(*b*) The stability of the measuring device [the potential-divider ratio and battery voltage (V_R)] must be within less than 1 in S. Noise and drift of the feedback amplifier, referred to its input, can be regarded as fluctuations in the reference potential.

(*c*) The feedback circuit must be free from self-oscillations.

Consider an arrangement of this type designed to stabilize a 50-kV power unit in the presence of initial fluctuations of 0·5% such as may occur from the mains supply after passage through a constant-voltage transformer. A reasonable potential-divider ratio would be 200 : 1, i.e. a reference voltage of 250 volts. For a stability within 2 parts in 10^6, the following conditions would have to be met:

Amplifier gain × modulator gain × potential-divider ratio = 2 500.
Amplifier gain × modulator gain = 500 000.
Potential-divider ratio and reference-voltage stability = 2 parts in 10^6.
Amplifier input stability = 0·25 mV.

The last two requirements present very great practical difficulties. Battery stabilities within a few parts in 10^6 are possible.[6,7] Amplifier input stabilities within 0·25 mV are difficult to obtain over long periods, particularly with high input impedances which preclude the use of contact-modulator type amplifiers.[8] The conditions could be somewhat improved by using a lower potential-divider ratio and a reference battery of high voltage, but a really useful saving cannot be made, except by using a battery voltage so high that it becomes impracticable. In any case, the requirements for a potential-divider ratio constant to 2 parts in 10^6 are likely to be extremely difficult to meet. Whilst this is not considered impossible, it is doubtful whether such a figure could be economically obtained.

For these reasons, other possible devices to replace the potential divider were considered. The various possibilities reviewed will not all be discussed in detail; they include electrostatic weighing, generating-voltmeter and electron-optical methods. Finally, an electron-optical method was chosen, and it is the realization of this method which is discussed in the paper. The method is shown to have the advantage over a resistance potential-divider of increasing the signal level at the feedback-amplifier input from a fraction of a millivolt to a fraction of a volt at least.

The third main requirement for the feedback stabilizer circuit has not so far been discussed, but will be referred to in Section 5.3. It is, in practice, comparatively simple to achieve freedom from self-oscillation of the feedback loop.

(3) THE ELECTRON-VELOCITY ANALYSING METHOD

The principle of the stabilizing method used is to accelerate a beam of electrons through the voltage to be stabilized, to measure the velocity by deflection in a fixed magnetic field, and so to obtain a signal to the feedback amplifier. Similar methods have been described before,[9,10] but only for stabilities within a few parts in 10 000.

There are many possible methods of analysing or measuring the velocity or velocity distribution of a stream of charged particles. The one chosen, the 180° magnetic velocity analyser, is not the best, since it suffers from electron-optical aberrations which are avoided in others, but it was considered the most convenient for the present work, and other systems are not discussed in the paper.

The electron beam is directed into a uniform magnetic field (Fig. 2), in a direction perpendicular to the field. It is well known that the beam is then deflected in the field in the arc of a circle. If the magnetic field is limited at a boundary AB perpendicular

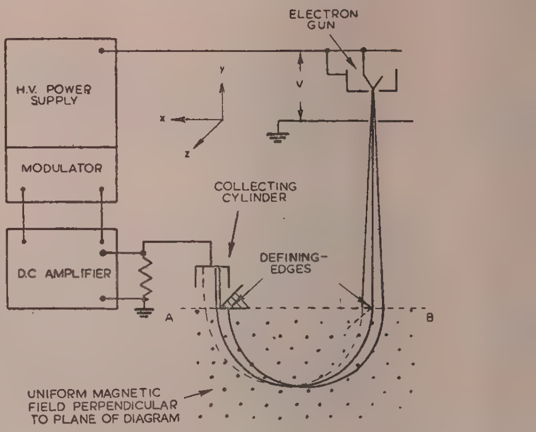


Fig. 2.—Block diagram of velocity-analyser voltage-stabilizing circuit.

to the direction of the incoming beam, the beam emerges from the field in a direction opposite to that of incidence.

It will be seen later that, in order to detect a fractional change in voltage of 1/500 000, changes in the position of the beam at the output must be detected of the order of 1/1 000 000 of the radius of the electron path. In a conventional type of velocity analyser such measurements are made by defining the width of the beam at the analyser input with a slit system. This is necessary in a true spectrometer, which serves the purpose of measuring a distribution of energies or masses analysed by the action of the deflection field. In the present instance, only one mass and one energy are present at a given instant. However, a slit system could still be used and has been considered, but the rather simpler arrangement illustrated in Fig. 2 is preferred. The beam from the electron gun, after deflection in the analyser field, falls on an earthed defining-edge which cuts off most of the beam and allows only a small fraction to pass into a collector cylinder. The current from this cylinder is passed through the grid resistor in the input stage of the feedback amplifier. It will be seen that an increase in voltage on the electron gun increases the beam orbit diameter, and hence the collected current, thus producing a negative signal to the amplifier.

For the sake of the following discussion, three rectangular co-ordinates, x, y and z, are defined, the y-direction parallel to the incoming electron beam, the x-direction parallel to AB, and the z-direction that of the magnetic field. Fringing being neglected, the diameter of the electron orbit in the magnetic field (x_D) is given by

$$x_D = \frac{6 \cdot 7\sqrt{V}}{B}\ \text{cm} \qquad (1)$$

where V is the voltage across the electron gun (and that to be stabilized) in volts, and B is the flux density in the uniform magnetic field in gauss.

The change in the beam orbit diameter for a small voltage change δV is

$$\delta x = \tfrac{1}{2} x_D \delta V / V \qquad (2)$$

(3.1) Effect of Variations in the Electron-Gun Current

If the beam overlaps the collector defining-edge by a distance δx corresponding to a voltage change $\delta V'$, a fractional change, i, in the total beam current from the electron gun will result in a spurious voltage change of $i\delta V'$. Clearly, this spurious change must be less than V/S, where $1/S$ is the fractional stability required.

Thus,
$$i\delta V' < V/S$$
$$\delta V' < V/iS$$

If the sensitivity, g, of the analyser is defined as the change in collector current for a 1-volt change in the high voltage, maximum allowable values for the collected current (δI) and overlap distance (δx_D) are given by

$$\delta I' = g\delta V' = gV/iS \quad . \quad . \quad . \quad . \quad . \quad (3)$$

and

$$\delta x_D = \tfrac{1}{2}x_D/iS \quad . \quad . \quad . \quad . \quad . \quad . \quad (3a)$$

These equations are discussed below in other connections, but at the moment it should be noted that, since $\delta I'$ must be small, only the edge of the electron beam must be allowed to spill over to the collector cylinder. This leads to a difficulty, because the electron beam from the electron gun is not well defined at the edges, but has an approximately Gaussian distribution of current density.[12] Thus, if the edge only is effectively used, the low current-density leads to a low sensitivity, and hence to low signal output and an inefficient utilization of the electron beam. For this reason, a second defining-edge is placed at the analyser input. This edge is arranged to bisect the beam so that the edge of the beam spilling over the second defining-edge is sharp and of maximum current density. The input defining-edge and the collimation of the beam are shown dotted in Fig. 2. The input defining-edge serves another important purpose by defining the effective position of the beam at the analyser input. Thus, small drifts in position of the beam, due, for example, to the gun filament moving, produce only second-order disturbances in the system.

(3.2) The Analyser Aberrations

At the output of the 180° analyser, a focused image in the x-direction of the current-density distribution at the input is produced. In the z-direction, the beam is not focused and therefore spreads at the same rate as before entering the field. Since the output defining-edge is extended in the z-direction, this spreading is of no particular significance in the following argument.

The analyser does not produce a perfectly sharp image but suffers from radial aberration dependent on the beam angle of the illuminating cone. The effect of this aberration is now considered. In Fig. 3, YO is the direction of the axis of the cone

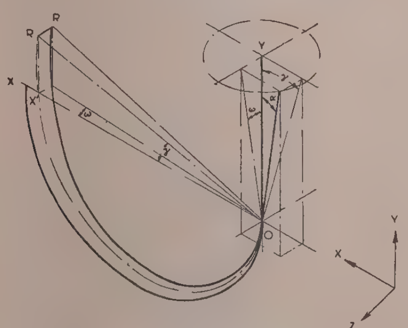


Fig. 3.—Diagram illustrating the derivation of the analyser aberration.

of electrons entering the spectrometer; OX lies along the boundary of the magnetic field and is equal in length to the diameter (x_D) of the orbit of the central electron ray. A marginal ray, making an angle α with the axis of the illuminating cone, may make angles of ω with the XY plane and of γ with the YZ plane.

Then $\tan^2\alpha = \tan^2\omega + \tan^2\gamma$.

The marginal electrons will move in a trajectory in the magnetic field with a constant velocity $v \sin\omega$ parallel to the field, and the projection of the trajectory in the plane YOX will be a circle of diameter OR′ ($= x_D\cos\omega$) making an angle γ with OX; v is here the full velocity of the electrons. The projected trajectory will finally cut OX in X′. XX′ is the aberration of the system.

Now XX′ = OX − OX′ $= x_D(1 - \cos\gamma\cos\omega)$. The maximum value of XX′ occurs when $\cos\gamma\cos\omega$ is a minimum. This can be shown to occur when $\tan\gamma = \tan\omega = 1/\sqrt{2}\tan\alpha$.

Hence $$XX'_{max} = x_D\left(1 - \frac{1}{1 + \frac{1}{2}\tan^2\alpha}\right)$$

or for small values of α

$$XX'_{max} = \tfrac{1}{2}x_D\alpha^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Thus, the sharpness of the image of the defining-edge can be improved to any desired degree by reducing the illuminating angle. This can only be done, however, at the expense of reduction of current density in the spot. The beam-spread parallel to the magnetic field due to the component of velocity in that direction amounts to $\pi x_D\alpha$, and is of consequence only in limiting the spacing of the deflecting magnetic pole-pieces.

As a result of the aberration the edge of the beam is not sharp and the collected current cannot be reduced below a certain minimum if the maximum sensitivity is to be maintained. Alternatively, to reduce the collected current to a prescribed minimum without loss of sensitivity due to aberration, the aperture angle of the beam entering the analyser must be limited.

This condition leads to a maximum allowable aperture angle. If the aberration is specified as being less than half the overlap distance [δx_D in eqn. (3a)], the maximum allowable value of α is given by combining eqns. (3a) and (4) as

$$\alpha^2 \leqslant \frac{1}{2iS} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

For example, if $i = 10^{-2}$ (1% beam current stability) and $S = 10^6$ (1 in 10^6 h.v. stability), α must be equal to or less than 7×10^{-3} rad.

(3.3) Beam Intensity and Sensitivity

If the electron beam is bisected by the first defining-edge, the current collected by the second edge for an overlap x is given, to a close approximation, by

$$I = 4\cdot 5J_0r_0^2\left[1 - \exp\left(1\cdot 3\frac{x}{r_0}\right)\right] \quad . \quad . \quad . \quad (6)$$

which expression is derived in the Appendix, Section 10.

Differentiating, we have

$$\frac{dI}{dx} = 5\cdot 9J_0r_0\exp\left(-1\cdot 3\frac{x}{r_0}\right)$$

Eqn. (2) gives $\delta x = \frac{1}{2}x_D\delta V/V$, and hence the sensitivity is given by

$$\frac{dI}{dV} = g = \frac{2\cdot 9x_DJ_0r_0}{V}\exp\left(-1\cdot 3\frac{x}{r_0}\right)$$

For a small overlap distance (i.e. $x \ll r_0$) $\exp(-1\cdot 3x/r_0)$ approaches 1 and hence

$$g = \frac{2\cdot 9x_DJ_0r_0}{V} \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

The maximum current density per unit solid angle (brightness β) obtainable from an electron gun is limited, as has been shown from theoretical considerations by Langmuir.[11] The design and operating conditions necessary to obtain the

theoretical brightness have been discussed by Haine and Einstein.[12] The theoretical maximum value is given by

$$\beta = \frac{J_c V}{\pi k T}$$

Thus, the maximum current density at the first defining-edge is given by

$$J_0 = \beta \pi \alpha_{max}^2 = \frac{J_c V \alpha_{max}^2}{kT} \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where α_{max} is the maximum beam angle, which must not exceed that defined by eqn. (5).

The theoretical sensitivity of the analyser can now be enumerated after deciding on suitable values for x_D and r_0. From general considerations, x_D was chosen to be 15 cm. If the electron gun is to be used without a condensing lens, it should be placed at such a distance from the input defining-edge that the effective source subtends an angle $2\alpha_{max}$ at the edge. The radius of the illuminating spot (r_0) is given by the product of the distance from gun to spectrometer input, and the semi-angle of the beam leaving the electron gun. This angle can be controlled by variation of the electron-gun bias without appreciable variation of the brightness β (see Reference 12). The most relevant practical factor affected is the total current in the electron beam, which it was desirable for practical reasons to keep below 100 μA. This current is given by

$$I = \beta \times \pi \alpha_{max}^2 \times \pi r_0^2$$

Hence $$r_0 = (IkT/J_c V \pi \alpha^2)^{\frac{1}{2}} \quad . \quad . \quad . \quad . \quad . \quad (9)$$

It is now possible to evaluate the sensitivity in terms of practical values of the various parameters. These are:

$$\alpha_{max} = 7 \times 10^{-3}$$
$$x_D = 15 \text{ cm}$$
$$J_c = 1 \text{ amp/cm}^2$$
$$T = 2\,800° \text{ K}$$

Hence, from eqn. (8) $J_0 = 10{\cdot}0$ amp/cm^2

and from eqn. (7) $g = 8{\cdot}7 \times 10^{-3}\, r_0$ amp/volt

By combining with eqn. (9) we obtain

$$g = 1{\cdot}5 \times 10^{-3}\sqrt{I} \text{ amp/volt}$$

e.g. for $I = 50 \times 10^{-6}$ amp
$g = 10 \times 10^{-6}$ amp/volt.

In combination with an output load resistance of 10^6 ohms, the analyser should thus be capable of giving a gain of at least 10 times. This demonstrates one important advantage over a resistance potential-divider where the signal level at the tap point corresponding to the maximum permissible variation in the high voltage is of the order of a millivolt or less, thus requiring a higher amplifier gain and a much more stable input level. The analyser should not require an amplifier input stability closer than ± 1 volt for an overall stability of 1 part in 10^6 with a high voltage of 100 kV.

In practice, to obtain a value of $\alpha = 7 \times 10^{-3}$ rad without the use of a condenser lens between the electron gun and spectrometer, it is necessary for the electron gun to be less than 4 mm from the analyser input. Since this is not practicable, and since it was thought desirable to avoid the use of a condenser lens, the gun was positioned 8 cm from the analyser input, giving an input angle of 3×10^{-4} rad. (The effective source size in the type of gun used is 0·0025 cm radius.) This reduction in angle results in a reduction in sensitivity from 10 μA/volt to 0·4 μA/volt. This figure still gives an adequately high signal level to the amplifier input, and the reduction in current density at the defining edges reduces the surface thermal load, which would be heavy under the conditions previously prescribed. In practice, the sensitivity usually obtained is found to be within a factor of 2 of this calculated value.

It was found advantageous to stabilize the electron-gun current to within less than 1% in order to work with a larger overlap current. This decision arose after difficulty was experienced in making the two defining-edges exactly parallel. It will be seen that this sets a lower limit to the overlap current that can be used without loss of sensitivity. It should not be impossible to overcome this mechanical difficulty, but the easier alternative of stabilizing the beam current was adopted in this instance.

There are still a number of factors which have not been discussed and which might be expected to provide limitations to the stability attainable with the electron-velocity analyser. These factors are enumerated and discussed in Sections 3.4 to 3.7.

(3.4) Electrical Effects

The spread in the velocity of emission of the electrons sets a limit to the final stability. The mean energy spread due to this cause is given by $T/11\,600$ electron volts, where T is the cathode temperature in deg K. For tungsten at 2 800° K the spread is about 0·25 volt, giving a stability of 1 part in 250 000. In practice, the electron gun can make the beam more monochromatic, as has been explained by Haine and Einstein.[12]

(3.5) Mechanical Vibration

Spurious changes will result if the two knife edges move relatively. For the conditions of the above example, the knife-edge separation must be stable within $x_D \times 10^{-6}$, or 0·15 micron for $x_D = 15$ cm. This presents some difficulty, but seems by no means impossible, particularly since the knife edges can be supported at either end of a rigid bar.

(3.6) Thermal Expansion

Thermal expansion will cause variations in the magnet field, and in the distance between the defining-edges. A typical temperature coefficient of 20 parts in 10^6 per deg C leads to a temperature-stability requirement of 1/20° C for a voltage stability of 2 parts in 10^6. This could be reduced by using materials of low coefficient of expansion or by compensation. As will be seen, an even smaller temperature tolerance is set by the temperature coefficient of the permanent magnet used to produce the deflecting field.

(3.7) Magnetic-Field Variation

Since the diameter of the electron orbit in the analyser is inversely proportional to the magnetic field, it is clear that the high voltage cannot be more stable than the magnetic field strength. Three possible methods of exciting the magnet gap have been considered: a permanent magnet, an electromagnet supplied from a stabilized current source, and an electromagnet connected in series with the electron-microscope objective lens. This last method has a further advantage which is discussed below.

Permanent-magnet excitation is attractive in that it eliminates the necessity for any associated circuits. Modern permanent-magnet alloys have temperature coefficients of about 1 part in 4 000 per deg C,[13] so that the temperature must be kept constant to 0·01° C to maintain the field constant to 1 part in 500 000. To minimize drift in the strength of the magnet, it should be over-excited and then stabilized by the application of an a.c. field.

The possibility of compensating the magnet against variation due to temperature changes, by using shunts of magnetic material

whose permeability is temperature-sensitive,[14] has been considered but has not so far been applied.

An electromagnet requires a current supply stable to 1 part in 500 000, although it is possible to operate it in series with the electron-microscope objective lens, in which case the change in focal length due to a current change in the lens would be compensated by the high-voltage change resulting from the same change in current occurring in the analyser magnet. This compensation would be exact only where iron saturations, if present, were identical in both magnet and lens. It is difficult to achieve these conditions in practice.

A further possible cause of variation in the deflector field lies in the varying stray fields arising from, e.g., nearby transformers. The maximum permissible stray field for a total deflector field of 100 oersteds is 0·2 mG. This is roughly equal to the stray field measured in the laboratory, where precautions had already been taken to minimize stray fields because of interference with the electron microscope itself. An advantage of the electromagnet is that it provides a large degree of shielding against stray magnetic fields. The permanent-magnet arrangement is poor in this respect because of the low permeability of the permanent-magnet material under operating conditions.

(4) THE PRACTICAL REALIZATION OF THE VELOCITY ANALYSER

(4.1) Velocity Analyser

The analyser system finally used is shown in Fig. 4. It consists of an electron gun (1) similar to that developed for an electron microscope,[15] using a 0·005 in tungsten wire hairpin as cathode, a biased cathode shield, and earthed anode plate. The anode plate can be adjusted laterally to centre the electron beam as described in Reference 12, which deals comprehensively with the performance and characteristics of this type of electron gun.

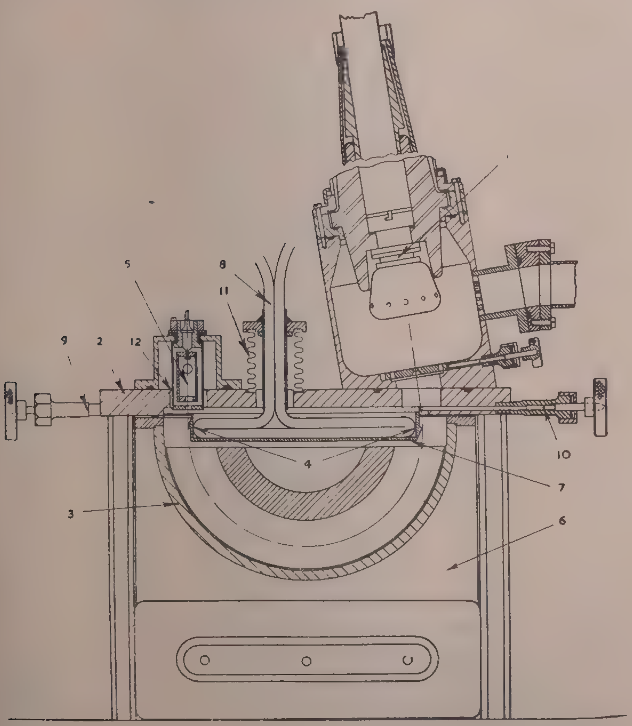


Fig. 4.—Cross-section of 180° magnetic velocity analyser.

Distance between vertical support rods = 25 cm.

The electron gun is mounted directly onto the 180° analyser top plate (2), its axis making an angle of 6° to the top edge of the magnet. The tilt of the gun is desirable because the electron beam is deflected by the fringe field of the magnet before passing into the region between the magnet pole-pieces. For the spectrometer aberrations to have a minimum effect, the electron-beam axis should be perpendicular to the line joining the input and output defining-edges of the spectrometer. The angle of 6° was computed from the magnetic-field distribution measured on a resistance-network analogue.[16]

The analyser itself consists of the semicircular brass vacuum vessel (3) with rectangular top flange, the defining-edges (4), the collector cylinder (5), and the magnet (6). The vacuum vessel has a copper liner which can be removed for cleaning. It is important, as in most electron-optical systems, to avoid contamination by insulating layers which may become charged by stray electrons to a potential sufficient to deflect the main electron beam.

The defining-edges are machined in a single brass bar (7) with copper inserts soldered in at the ends. The actual edges are lapped straight to microscopic tolerance and tapered on the back to allow the electron beam to spread without interception. It is important that the edges should be accurately parallel, as has already been mentioned. The copper blocks are fitted with an oil-cooling pipe (8) soldered into a hollow which is milled lengthwise. Adjusting screws (9,10) enable the first edge to be centred in the incoming electron beam by traversal in the x-direction. The oil connections are taken out through a flexible metal bellows (11).

The collector is a shielded and insulated cylinder, with an entry aperture arranged to collect the beam which spills over the second defining-edge and to trap as many back-scattered electrons as possible. It must be borne in mind that, at voltages of 50–100 kV, as much as 60–70% of the beam falling on a metal surface can be lost by back-scattering. A coating of graphite is applied to minimize this loss.

A separately insulated second apertured box (12) is fitted round the collector, and serves to indicate (by absence of any current reaching the electrode) correct alignment and adjustment of the beam.

The permanent magnet consists of two rectangular slabs of ½ in-thick mild steel measuring 9 in × 7 in, spaced 0·75 in apart. Excitation is produced by a solid cylinder of Alcomax II permanent-magnet material fitting between the pole-piece slabs. This arrangement represents a very efficient use of the permanent-magnet material since the drop in magnetomotive force along it is uniform and none of the material is wasted. It also results in a simple design for the permanent magnet, since no allowance for demagnetization by unknown leakage flux of the magnet itself need be made. The deflecting field for 100-kV electrons (the maximum voltage allowed for in the design) is 150 oersteds. This can be obtained with a 2 cm length of Alcomax II operating at a flux density of about 10 kG. The magnet area required is therefore about 9 cm^2. .

The alternative electromagnet comprises two Mumetal poles of the same size as those in the permanent-magnet system, and a Mumetal core of cross-section large enough to keep the flux density of the core below 5 000 G. The core is excited by a winding capable of giving a flux density up to 150 G in the gap. This magnet gives a current/flux-density relation which is linear to within 0·1%, and a flux density differing by less than 0·1% from that in another similar magnet, excited in series with it. This is of importance in one of the tests applied to the system and mentioned in Section 6.3.

(4.2) Temperature Stabilization

The magnet assembly is immersed in an oil-filled tank kept at a constant temperature of 28° C by a mercury-toluene regulator with thyratron relay controlling the heater. The heater is a 60-watt electric light bulb, which provides heating with a small 50-c/s magnetic field. The oil is stirred by small d.c. motors fitted 3 ft from the magnets. The temperature of the bath is stable within about $\pm 0 \cdot 01°$ C. A small gear pump attached to one of the stirrers pumps oil through the defining-edge cooling pipe.

(5) ELECTRICAL CIRCUITS

(5.1) Electron-Gun Supplies

Normally, the electron gun of the electron microscope operates with a bias obtained from the voltage drop in a resistance in series with the main high voltage (self-bias). In the stabilizing arrangement at present employed, it is desirable to connect the microscope and analyser cathodes directly together electrically, as the stabilizer effectively stabilizes the potential at its cathode, and any intervening circuit would be likely to introduce a varying potential on the microscope cathode relative to the analyser. On the other hand, it has been seen (Section 3.1) that the analyser beam current should be stable to better than 1%. It is difficult to achieve this stability without self-bias unless some form of current stabilizer is used. Fortunately, it was possible to devise a very simple circuit giving a current stability of about 1 part in 1 000.

Circuits for cathode heating, biasing and current stabilizing are mounting on a platform insulated for 100 kV from earth. The circuit arrangement is shown schematically in Fig. 5. Two

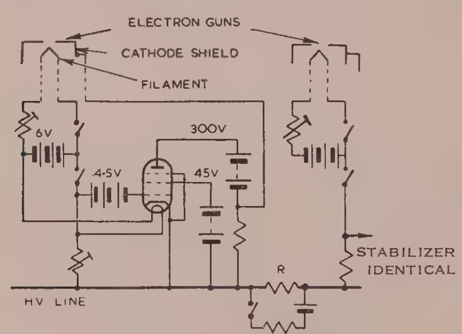


Fig. 5.—Circuit diagram of h.v. electron-gun current stabilizer.

identical circuits are included to feed two electron guns, although only one stabilizer circuit is shown in the Figure. Filament heating is obtained from 6-volt car-type accumulators. An on-off switch and a rheostat control are operated by means of long insulating rods. The current stabilizer comprises a single degenerative amplifier stage using a type 6F13 pentode valve. A small reference resistance, in series with the high-voltage supply, produces the control signal, which is fed to the grid of the valve through a 4·5-volt battery. The output of the amplifier provides the bias to the electron gun.

Provision is made for injecting a fixed voltage across the resistor R (Fig. 5) connected in the high-voltage line between the two electron guns. This voltage is used to calibrate monitoring equipment and to introduce accurately known fractional changes in the focus of the microscope.

(5.2) High-Voltage Supply

The high-voltage power unit consists of a voltage-doubling Greinacher circuit with a resonant air-cored transformer energized by a 25-kc/s oscillator. The arrangement is very similar to that used in the electron microscope described by Haine, Page and Garfitt.[15] The schematic circuit is included in Fig. 6. Variation of the output is effected by control of the

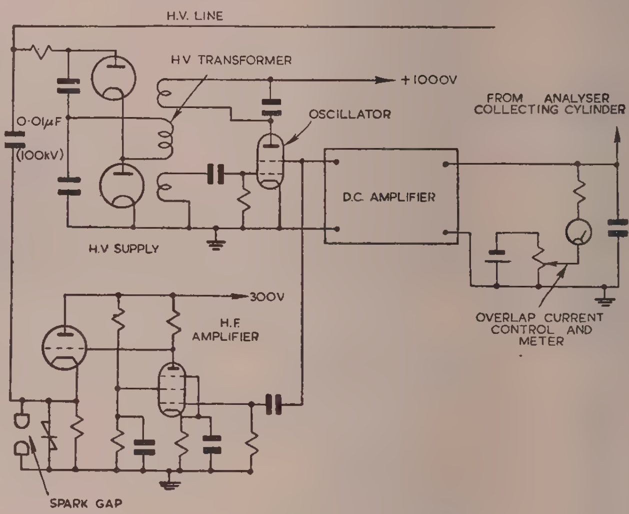


Fig. 6.—Circuit diagram of h.v. power unit with parallel d.c. and a.c. feedback amplifiers.

oscillator screen-grid voltage, the zero-frequency gain being about 250 from oscillator screen-grid to output high voltage. The gain at higher frequencies is also important and is discussed in Section 5.4.

(5.3) Feedback D.C. Amplifier

To maintain a stability of 2 parts in 10^6 in the presence of variations of input and load of 1%, an overall feedback gain of 5 000 is required. If the zero-frequency gain of the modulator is 250, and that of the analyser about unity with a 5-megohm load, an amplifier gain of at least 20 is desirable. The input to the amplifier must be stable to within 0·1 volt.

The amplifier is conventional in design except that care had to be taken to make the gain independent of frequency over the range 0–50 kc/s and that low-grid-current valves are used in the input stage. The output valve feeding the screen-grid of the oscillator stage is a cathode follower. A microammeter in series with the input resistor indicates the overlap current collected in the collecting cylinder; this current is adjusted by means of a variable direct voltage in series with the microammeter. When correctly adjusted, the high voltage recovers automatically after a large disturbance such as a flashover of the microscope gun. A reduction in high voltage causes the overlap current to fall to zero, so applying maximum screen voltage to the oscillator valves and re-establishing the stabilized voltage.

(5.4) Self-Oscillation and Frequency Response

Bode[17,18] has shown that, with certain reservations, the oscillation stability of a feedback amplifier can be judged from the log-gain/log-frequency characteristic. Instability results if the gain is greater than unity at the point where the negative slope of the characteristic exceeds 2; at this point the phase shift is 180°. The frequency response of the h.v. power supply was measured, and is shown in curve (*a*) of Fig. 7. A maximum stable gain of only 100 would be expected from the Bode criterion; this was approximately confirmed by experiment. Not only is the circuit unstable, but the gain at 50 c/s and harmonics of 50 c/s..

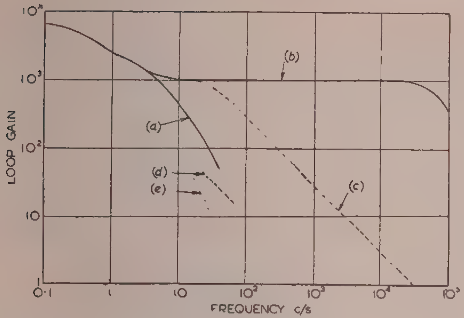


Fig. 7.—Loop-gain/frequency characteristics of feedback loop with and without the modification produced by a.c. feedback amplifier and stabilizing time-constant.

is too low to allow mains frequency and harmonics to be reduced appreciably by the stabilization. For larger gains to be applied, it is clear that the frequency response of the feedback loop had to be modified so as to make the gain fall off more slowly with increasing frequency. The parts of the circuit causing the rapid fall-off were the oscillator and smoothing circuit. Improvement was achieved by the use of an auxiliary a.c. feedback loop, by-passing the h.v. power unit from oscillator screen grid to the h.v. line. This amplifier consists of one pentode stage and a cathode follower connected from the output of the d.c. amplifier to the earth side of the final smoothing capacitor. A spark-gap and non-linear resistor are connected across the load resistor for protection against h.v. surges. The arrangement is shown in Fig. 6.

The improvement resulting from the insertion of this amplifier is shown in curve (*b*) of Fig. 7. The gain is now constant up to 50kc/s, but oscillations would still occur at about 200kc/s, where the slope of the characteristic reaches −2. This instability can be prevented by lengthening one of the time-constants to give the dotted curve (*c*) in Fig. 7, so reducing the gain to unity before the slope reaches −2. For comparison, slopes of −1 and −2 are shown by the short dotted curves (*d*) and (*e*). In practice, the large time-constant was applied in the form of a capacitor across the input of the d.c. amplifier input resistor. By adjustment of this capacitor and the a.c. amplifier gain control, a total stable loop gain of up to 5 000 was obtained. The insertion of the loading capacitor in this position has the incidental advantage of reducing the input impedance of the amplifier, and hence the possibility of pick-up of stray transient voltages.

(5.5) Cathode-Ray Oscillograph Monitor

The ripple voltage on the 50-kV line is monitored on a cathode-ray oscillograph and a 30 : 1 capacitance potential-divider formed from a 100cm section of the h.v. cable, which has its sheath isolated and connected to earth through a 3 000 $\mu\mu$F capacitor. The sheath is connected to the input of the oscillograph amplifier.

(5.6) Auxiliary Analyser

In addition to the main stabilizing analyser, a second identical analyser is available. The second analyser can be switched in place of the first as a standby in case of filament or other failure, or it can be used as a test monitor to check the h.v. stability. Its use in this connection is discussed in Section 6. The output of the second analyser is recorded on a 4in chart recorder via a d.c. amplifier. This arrangement has a sensitivity capable of showing a change in high voltage of 1 part in 10^6 per millimetre division of the chart. The stability of the amplifier corresponds to changes of less than 1 part in 10^6 in the high voltage.

(6) PERFORMANCE AND OPERATION OF THE STABILIZER SYSTEM

(6.1) Operation

The measurement of the stability of the output of a voltage stabilizer necessitates comparison with a device at least as stable as the stabilizer itself. It was not practicable, in the present instance, to assess the stability by the microscope performance, since so many other factors may affect it. The only alternative was to use a second analyser in parallel with the one used for stabilizing, and to measure the variation of its output. The measured variation will be expected to result from the sum of actual changes in the high voltage and variations introduced by the measuring analyser. In general, it is comparatively simple to produce a stabilizer which reduces the input and load variations to a prescribed level. On the other hand, difficulty usually arises because the stabilizing system produces its own variations in the output such as those due to drift in component values and valve characteristics, and to pick-up of stray voltages. The problem is to locate and eliminate all such sources of spurious voltage change.

The following test procedure was adopted, the general arrangement being shown in Fig. 8:

(*a*) The d.c. amplifier is run with no collected current at the input and its output voltage variations are measured.

(*b*) The stability of the two electron-gun currents is checked by adjusting the position of the analyser defining-edge block and the energizing field so that the whole of the gun current is received in the collector cylinder. The stability is checked by recording the output via the d.c. amplifier suitably backed-off to allow a reading sensitivity of 1 part in 1 000.

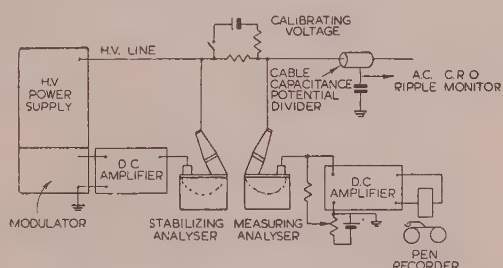


Fig. 8.—Block diagram to show test arrangement.

(*c*) The two defining-edge blocks are now moved to bisect the beam at the input as indicated by a reduction in collector current to 50%. The magnetic field is increased until the beam is intercepted by the output defining-edge and the collected current reduced to the amount specified in eqn. (3).

(*d*) The sensitivity of the two analysers is now measured with the aid of the calibrating voltage referred to in Section 5.1.

(*e*) The voltage variation developed at the output of the stabilizing analyser is now measured. This variation, divided by the known gain of the analyser, would give the variation on the h.v. line if the analyser itself were perfectly stable. In fact, it only gives the effective variation from all parts of the feedback loop other than the analyser.

(*f*) The complete system is now operated with ripple monitor and chart recorder in operation. The chart recorder is calibrated directly from the calibrating voltage between the two analysers.

(6.2) Results

Typical test results for the apparatus as described when first operated are enumerated below:

Beam current stability	1 in 10^3 (for 1 h)
Equivalent stability at stabilizing analyser output	1 in 10^7 (for 1 h)
Ripple voltage	2 in 10^5
Long-term drift	up to 10 volts/min.

The high equivalent stability at the analyser output indicates that adequate feedback gain is available, and that spurious fluctuations in the amplifier and modulator are negligible, so that any measured voltage fluctuations must result from instability in one or both of the analysers. The ripple voltage was excessive and was subsequently reduced by minor circuit changes to 4 parts in 10^6.

It was not possible to ascertain whether observed variations result from the stabilizing analyser or the measuring analyser, and it was therefore necessary to attempt to improve both. In the first attempt, the actual overall temperature-coefficients of the analysers were measured by causing large changes in the temperature of one oil bath and measuring the indicated change in high voltage. In this way, for the permanent magnets, a temperature coefficient of 1 part in 4 000 per deg C was measured, while for the electromagnets the measured value was 1 part in 20 000 per deg C. Since the temperature was found by measurement to remain stable to 0·01° C, voltage changes of not more than 1 part in 400 000 should result.

The main source of drift was found to be incorrect alignment of the analysers. In particular, it was found that over the first 10–15 min of operation the electron gun filaments tended to move, thus moving the electron spot at the analyser input. By delaying the final alignment until after such an initial period of operation, a considerable improvement was obtained. Finally, the drift was reduced to about 1 part in 10^6 per minute, with some day-to-day variation. The overall stability then obtained is shown in the records of typical runs in Fig. 9. The stability

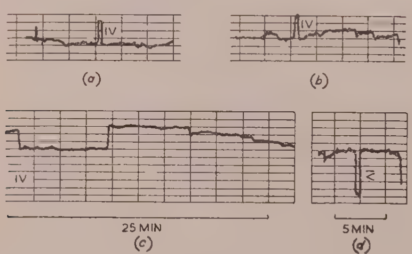


Fig. 9.—Typical test records from recorder with 1-volt calibration changes.

Chart scale in (*a*), (*b*) and (*c*) = 6 parts in 10^6 per vertical division.
Chart scale in (*d*) = 3 parts in 10^6 per vertical division.

can usually be relied upon to 2–3 parts in 10^6 for minute intervals. The ripple is still 4 parts in 10^6, but it is known that this can be reduced, and steps in this direction are being taken in a new h.v. power unit.

The above results were obtained with the electromagnets which were excited in series, so that the variations in excitation current were automatically compensated as previously described. A current stabilizer, giving a stability of 3–4 parts in 10^6 has now been developed by Jervis,[19] and is in operation feeding the microscope objective lens and the stabilizing analyser. Results with the permanent magnets have shown comparable results.

(7) CONCLUSIONS

A voltage stabilizing system has been constructed capable of stabilizing a h.v. d.c. supply to a few parts in 10^6 over most ten-minute periods. It is felt that stability over longer time intervals could be achieved but further experimental work is necessary to locate the causes of residual drift

(8) ACKNOWLEDGMENT

The authors wish to thank Dr. T. E. Allibone, F.R.S., Director of the A.E.I. Research Laboratory, for permission to publish the paper.

(9) REFERENCES

(1) HAINE, M. E.: "The Stabilization of High D.C. Voltages using an Electron Velocity Spectrometer," Comptes rendus du premier congrès international de microscopie électronique, Paris, September, 1950. *Revue d'Optique*, 1953.
(2) GABOR, D.: "Microscopy by Reconstructed Wavefronts," *Proceedings of the Royal Society of London*, 1949, **197,** p. 454.
(3) HANSON, A. O.: "Voltage Measuring and Control Equipment for Electrostatic Generators," *Review of Scientific Instruments, New York*, 1944, **15,** p. 57.
(4) HENKEL, R. L., and PETREE, B. "Stabilized High Voltage Power Supply for Electrostatic Analyzer," *ibid.*, 1949, **20,** p. 729.
(5) JERVIS, M. W.: "Direct Voltage Stabilizers in the Range 10–100 kV," *Journal of the British Institution of Radio Engineers*, 1954, **14,** p. 629.
(6) SENETT, W. P., and PIERCE, R. W.: "Exceptionally Stable Regulated Power Supply for Electrometer Tubes," *ibid.*, 1952, **23,** p. 534.
(7) JERVIS, M. W.: "The Temperature Coefficient of the E.M.F. of Kalium Cells," *Journal of Scientific Instruments*, 1953, **30,** p. 60.
(8) KANDIAH, K., and BROWN, D. E.: "High Gain D.C. Amplifiers," *Proceedings I.E.E.*, Paper No. 1196 M, November, 1951 (**99,** Part II, p. 314).
(9) VON ARDENNE, M.: "Elektronen Ubermikroskopie" (Julius Springer, 1950).
(10) STRAUSS, R.: "Ein neuer Quotientmesser für das Elektronenmikroskop," *Archiv für Elektrotechnik*, 1950, **40,** p. 49.
(11) LANGMUIR, D. B.: "Theoretical Limitations of Cathode-Ray Tubes," *Proceedings of the Institute of Radio Engineers*, 1937, **25,** p. 977.
(12) HAINE, M. E., and EINSTEIN, P. A.: "Characteristics of the Hot-Cathode Electron Microscope Gun," *British Journal of Applied Physics*, 1952, **3,** p. 40.
(13) BOZORTH, R. M.: "Ferromagnetism" (Van Nostrand, 1951), Chapter 9, p. 357.
(14) HEDDLE, T. A.: "On the Temperature Sensitivity of Special Magnetic Materials," *British Journal of Applied Physics*, 1953, **4,** p. 161.
(15) HAINE, M. E., PAGE, R. S., and GARFITT, R. G.: "A Three-Stage Electron Microscope with Stereographic Dark Field, and Electron Diffraction Capabilities," *Journal of Applied Physics, New York*, 1950, **21,** p. 173.
(16) LIEBMANN, G.: "Solution of Partial Differential Equations with a Resistance-Network Analogue," *British Journal of Applied Physics*, 1950, **1,** p. 92.
(17) BODE, H. W.: "Network Analysis and Feedback Amplifier Design" (Van Nostrand, 1945).
(18) BODE, H. W.: "Relations between Attenuation and Phase in Feedback Amplifier Design," *Bell System Technical Journal*, 1940, **19,** p. 421.
(19) JERVIS, M. W.: "A Precision Direct-Current Stabilizer" (see page 274).

(10) APPENDIX

Current Overlapping an Edge Collected by Half Plane intersecting Half Gaussian Electron Spot Image

It is clear that for the sake of this argument the spot may be considered as circular at the spectrometer output. The fact that it is drawn out parallel to the edge of the electrode does not alter the variation of collected current with voltage.

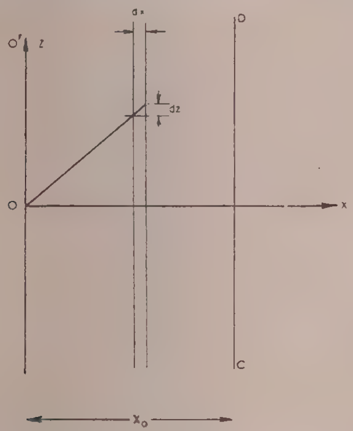


Fig. 10.—Illustrating calculation of current collected by plane intersecting electron-spot image.

Consider the case of a spot with a Gaussian distribution of half-width* r_0, total current I_t incident in the plane of the defining edge and with angular aperture α.

* Width at half-height.

In Fig. 10, O is the centre of the spot and CD is the defining edge parallel to the diametrical edge of the spot OO′.

Rectangular co-ordinates, z parallel to CD, and x perpendicular to CD, and passing through O, are introduced, and a small area dx, dz at a position x,z is considered.

The current density at a distance r from O is given by

$$J = J_0 \exp\left[-0\cdot7\left(\frac{r}{r_0}\right)^2\right]$$

where J_0 = current density at centre of spot, and
r_0 = half-width of Gaussian distribution curve.

$$dI = J_0 \exp\left[-0\cdot7\left(\frac{x^2+z^2}{r_0^2}\right)\right]dx,\, dz$$

The current overlapping the beam x_0 is

$$I = 2J_0\int_0^\infty\int_0^{x_0} \exp\left[-0\cdot7\left(\frac{x^2+z^2}{r_0^2}\right)\right]dx,\, dz$$

$$= 2J_0\int_0^\infty \exp\left(-0\cdot7\frac{z^2}{r_0^2}\right)dz\int_0^{x_0}\exp\left(-0\cdot7\frac{x^2}{r_0^2}\right)dx$$

The integrals are the probability integrals; the first has the solution $\sqrt{x/2} \times r_0/\sqrt{0\cdot7}$, while the exact value of the second can be obtained from tables. An approximate solution (to 1 or 2%) derived by curve matching gives

$$\sqrt{\pi/2} \times r_0/\sqrt{0\cdot7}\left[1 - \exp - \left(1\cdot6\frac{\sqrt{0\cdot7}}{r_0}x_0\right)\right]$$

Therefore $\quad I = 4\cdot5J_0r_0^2[1 - \exp(-1\cdot3x_0/r_0)]$

[The discussion on the above paper will be found on page 277.]

621.316.721

Paper No. 1730 M
Oct. 1954

A PRECISION DIRECT-CURRENT STABILIZER

By M. W. JERVIS, M.Sc.Tech., Associate Member.

(*The paper was first received* 3rd *March, and in revised form* 1st *July*, 1954. *It was published in October*, 1954, *and was read before the* MEASUREMENTS SECTION 14*th December*, 1954.)

SUMMARY

A d.c. stabilizer is described which delivers 50–150 mA stable to a few parts in 10^6. A degenerative feedback system is used with a contact-modulator amplifier, self-oscillation being prevented and the frequency response improved by an auxiliary a.c. feedback loop. Facilities for monitoring the stability are provided.

LIST OF SYMBOLS

R = Reference resistance, ohms.
R_L = Resistance of load, ohms.
V_R = Reference potential, volts.
I = Load current, amp.
V_i = Input voltage, volts.
μ = Amplification factor of series valve.
A = Gain of d.c. amplifier.

(1) INTRODUCTION

A previous paper has described the design and construction of an h.v. stabilizer[1] giving a stability within a few parts in a million in a 50-kV supply to an electron microscope. A similar order of stability is required for the magnetic deflecting field of the stabilizer and the magnetic objective lens of the microscope.

Magnetic fields can be stabilized by feedback arrangements from a search coil and fluxmeter;[2] the feedback may also be provided from a nuclear resonance probe,[3] which is capable of giving stability to a few parts in a million.[4] These methods are not applicable to the voltage-stabilizer electromagnet, since the field is too weak and the microscope-lens magnetic field is non-uniform and very limited in extent.

It appeared that the required field stability could be most easily obtained by exciting the lens and electromagnet with stable currents. The development and performance of a suitable current stabilizer along these lines is described in the paper.

(2) SPECIFICATION

The stabilizer was required to deliver a maximum current of 150 mA into a 3 000-ohm inductive load with a stability within 4 parts in 10^6 for periods of up to one hour. The output current was to be variable in steps of less than 1 part in 10^5 over a range of 10%, and the operating range preset by changing soldered joints or terminals to cover the total range of 50–150 mA.

(3) THE DEGENERATIVE FEEDBACK CURRENT STABILIZER

(3.1) Principle of Operation

Many forms of current stabilizers have been described,[5] but the most suitable for the load and stability requirements in question is the degenerative feedback type, illustrated in Fig. 1. A reference resistor R is connected in series with the load R_L and the voltage drop across R is opposed by a reference voltage V_R. The difference is applied to the input of a d.c. amplifier, the output of which feeds the grid of the series valve with the load R_L in its cathode circuit.

Mr. Jervis is in the Research Laboratory of Associated Electrical Industries, Ltd.

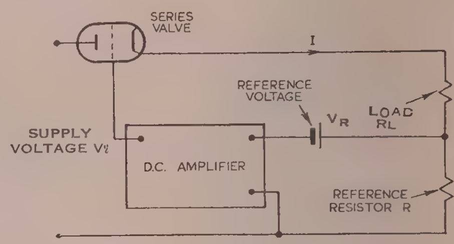


Fig. 1.—Degenerative feedback current stabilizer circuit.

The load current I is then given approximately by

$$I = V_R/R \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

It is well known[5] that in such a circuit the fractional load-current variation, $\delta I/I$, resulting from a fractional input voltage variation, $\delta V/V$, is given to a close approximation by

$$\delta I/I = \frac{R_L}{AR}\delta V/V \quad . \quad . \quad . \quad . \quad . \quad (2)$$

In addition, the fractional current variation, resulting from a fractional load-resistance change, $\delta R_L/R_L$, is given by

$$\delta I/I = \frac{R_L}{AR}\,\frac{1+\mu}{\mu}\,\frac{\delta R_L}{R_L} \quad . \quad . \quad . \quad . \quad (3)$$

if A is large.

It is not difficult to design a suitable amplifier to meet the requirements specified in eqns. (2) and (3), provided due care is taken to avoid feedback oscillation as discussed in Section 3.4. Important difficulties occur, however, in preventing spurious variations in the output current arising from sources within the feedback circuit.

(3.2) Reference Elements

It can be seen from eqn. (1) that the output current is directly dependent on the values of reference resistance and voltage, so that these must be stable to the same degree as that required in the output current.

It has been shown[6] that mercury primary cells have a temperature coefficient as low as 20 parts in 10^6 per deg C, so that temperature control to 0·05° C is adequate to reduce the effect of temperature to 1 part in 10^6. Similar temperature coefficients have been reported for other mercury cells.[7] Leclanché-type dry cells have a temperature coefficient of 300 parts in 10^6 per deg C.

The measurements described in Reference 6 showed that random fluctuations of about 5 parts in 10^6 occurred in the e.m.f. of the batteries tested. These were fitted with screw terminals. Later samples with copper-wire terminal connections gave random fluctuations of about 1 part in 10^6, and this type was chosen for use in the stabilizer.

The temperature coefficient of manganin resistance wire is less than 5 parts in 10^6 per deg C at 26° C. It can therefore be assumed that a constancy of 1 part in 10^6 over a few hours can be obtained in a reference resistor wound with wire, provided

its temperature is adequately stabilized and contact potential changes are avoided.

(3.3) Feedback Amplifier

Any internal fluctuations in the d.c. amplifier when referred to the input can be regarded as a voltage in series with its input. They therefore have the same effect as a fluctuation of the reference voltage, and must be less than the maximum permissible fluctuations in the reference potential.

Thermionic-valve d.c. amplifiers suffer from changes in effective grid potential of the input valve due to internal changes and supply-voltage fluctuations. In simple amplifiers, effective input variations of 10 mV/h are common. These can be reduced to 0·1–1 mV/h by careful balancing and selection of valves and suitable stabilization of the supply voltages.[8] Even under these conditions, the stabilizer would require a reference potential of at least 100 volts if the effect of these fluctuations is to be less than 1 part in 10^6. The use of such a high reference potential is inconvenient if the battery has to be temperature stabilized, and leads to a considerable power dissipation in the reference resistor.

This difficulty can be overcome by the use of either a galvanometer amplifier or a modulator-type amplifier. A relatively simple galvanometer amplifier can be made with an input stability to less than 1μV, and the use of such an amplifier in a current stabilizer has been described by Lawson and Tyler,[9] who obtained a stability of 20 parts in 10^6 for several hours. This type of amplifier is somewhat sensitive to vibration, and being fragile is easily damaged by overload.

With the contact-modulator amplifier, comparable stability is obtainable with a more robust instrument, not sensitive to vibration or easily damaged by electrical overload. Such amplifiers are commercially available in a convenient form,[10] and one of these was chosen for use in the stabilizer. In this particular instrument, the d.c. input signal is "chopped" into square waves by a vibrating contact. The resultant a.c. signal is passed through a step-up transformer, amplified in an a.c. valve amplifier and then rectified by a phase-sensitive detector. This type of amplification uses the valves as a.c. amplifiers only, so that changes in the effective grid potential of the valves merely alter the gain slightly. Thus, for a stability to 1 part in 10^6, a reference potential of 1 volt is adequate. A 5·2-volt battery was actually used, consisting of four mercury cells of the type described in Reference 6.

Negative voltage feedback is applied to the amplifier to increase its input resistance. On the sensitive range, the gain is $2{\cdot}5 \times 10^5$ with an input resistance of more than 1 000 ohms.

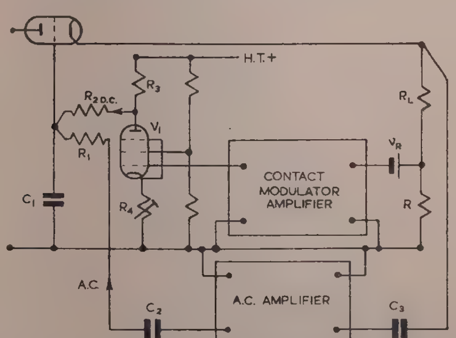


Fig. 2.—Block diagram of the precision direct-current stabilizer circuit.

$R_1R_2 = 1$ megohm $\quad C_1 = 0{\cdot}25\ \mu F$
$R_3 = 33\ k\Omega \quad C_2 = 0{\cdot}5\ \mu F$
$R_4 = 5\ k\Omega \quad C_3 = 0{\cdot}5\ \mu F$

For a stability to 1 part in 10^6, with input voltage variations of 1% and load variations of 5%, eqns. (2) and (3) give the minimum values of gain as $1{\cdot}2 \times 10^5$ and $3{\cdot}6 \times 10^6$ respectively ($R_L = 3\ 000$ ohms, $R = 50$ ohms, $\mu = 5$). The contact-modulator amplifier gain is $2{\cdot}5 \times 10^5$. This is followed by a single-stage amplifier V1, providing phase reversal and a gain of 10. The arrangement is shown in the block diagram, Fig. 2. The a.c. amplifier shown is referred to in Section 3.4.

(3.4) Loop Gain and Frequency Response

The contact-modulator amplifier has two disadvantages. First, the final rectification of the amplified chopped d.c. signal leaves some ripple in the output, consisting of the fundamental and harmonics of the chopping frequency. Secondly, the low-pass input and rectification filters cause the gain to fall rapidly

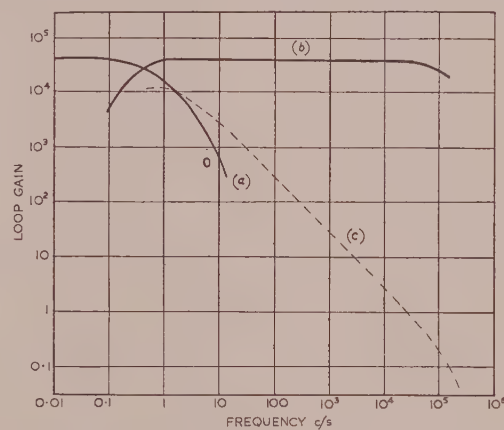


Fig. 3.—Frequency response curves of the feedback loop.

(a) D.C. loop alone.
(b) H.F. loop alone.
(c) H.F. loop with integrating circuit.

with increase of frequency. Curve (a) in Fig. 3 shows the measured frequency-response characteristic of the amplifier used. The high slope implies a large phase shift[11,12] which reaches 180° (slope, −2) at the point 0. A simple feedback arrangement incorporating this amplifier would therefore oscillate if the loop gain exceeded unity at 0, i.e. a d.c. gain of about 100. Furthermore, the loop gain at ripple frequency would be so small that the residual ripple introduced by the d.c. amplifier would not be appreciably reduced by the feedback.

The effective gain at high frequencies was increased by an a.c. amplifier connected in parallel with the d.c. path, as shown in the block diagram, Fig. 2. This technique has been described for use in servo systems by Williams *et al.*,[13] and Buckerfield[14] has shown how it can be used to extend the frequency range of contact-modulator amplifiers. Separate d.c. and a.c. loops have also been described for use in current stabilizers[15] and voltage stabilizers.[16]

The a.c. amplifier is connected to the high-potential end of the load to reduce the gain required. This is permissible, since no high-frequency variation in the load resistance is likely. To extend the response of the contact-modulator amplifier without interruption, the low-frequency response of the a.c. amplifier must be good down to 1 c/s. A resistance-capacitance coupled amplifier was used initially, but was found to be unsatisfactory owing to the long recovery time after blocking. It was therefore replaced by three stages of direct-coupled pentode amplification, with capacitance coupling at output and input only. The output of the a.c. amplifier is mixed with that of the d.c. amplifier

at the grid of the series valve. The frequency response of the a.c. amplifier is shown by curve (*b*) in Fig. 3. By suitable adjustment of the a.c. amplifier gain, the frequency response of the combined feedback amplifiers was made substantially constant from zero to 50 kc/s. The slope of this curve now reaches -2 at a frequency above 100 kc/s, but the amplifiers would still oscillate with a comparatively small loop gain. By inserting one long integrating time-constant (R_1 and C_1 in Fig. 2) at the grid of the series valve, the frequency response is modified to that shown by curve (*c*) in Fig. 3. The slope is now always less than -2 for gains above unity, and the circuit is stable for a d.c. loop gain up to 40 000 at least. A useful gain up to a frequency of 1 000 c/s is obtained.

The ripple voltage originating in the contact-modulator amplifier is about 100 mV peak-to-peak. After passing through the d.c. amplifying valve V1, the ripple is reduced by the filter R_2C_1 to about 60 mV. The ripple appearing across the load is further reduced by the feedback, the loop gain at 100 c/s being over 100.

(4) POWER SUPPLY AND TEMPERATURE CONTROL

The stabilizer is fed from a conventional full-wave-rectified supply operated from the 50-c/s mains. The power supply is followed by a relatively simple voltage stabilizer giving an output stable within about $\pm 0 \cdot 5\%$. This unit is desirable since it gives good stabilization against high-frequency disturbances, reduces ripple and prevents decoupling difficulties in the a.c. feedback amplifier described in Section 3.4.

The reference resistor and battery are enclosed in gasket-sealed brass boxes immersed in a water bath, temperature-stabilized to within $\pm 0 \cdot 05°$ C. A mercury-toluene thermostat, thyratron relay and heater element are used for temperature control.

(5) CONTROL CIRCUITS

The reference resistor is arranged to give 100 mA, variable over a range of approximately $\pm 5\%$. Other stabilized currents in the range 50–100 mA can be obtained by altering the reference voltage or the main reference resistor by means of soldered connections.

The main reference resistor is of 50 ohms, and is wound with manganin wire on a brass former which is sealed and immersed in the constant-temperature water bath kept at 26° C. To prevent the negative lead from forming part of this resistor, the lead is taken inside the resistor box directly to a potential terminal, as shown in Fig. 4. The positive connecting lead is made of thick copper to reduce changes in resistance arising from changes in room temperature.

Fine control of the current is obtained by a modified Kelvin–Varley slide system,[17] the first stage being 10 steps of 0·5-ohm manganin-wire resistors connected in series with the main reference resistor. Two switch tapping-points, spaced two steps apart feed a stage of 10 steps of 2-ohm nichrome resistors. The last stage is a simple rotary variable potential-divider of total value 100 ohms. It is not found necessary to stabilize these control resistors by thermostat since they form a small part of the total, but they are screened from rapid temperature fluctuations. The second resistor section is made higher than the first to reduce the effect of the contact resistance of the first switch.

This arrangement gives a control of the current to within 1 part in 10^4, although the steps are not accurately known. Small known changes in the output current are sometimes required and can be made by altering the reference voltage. For changes of ± 200 parts in 10^6, a voltage of ± 1 mV is obtained by passing a current of $\pm 100\ \mu A$ through a resistor of 10 ohms. This resistor is connected in series with the reference battery, as shown in Fig. 4.

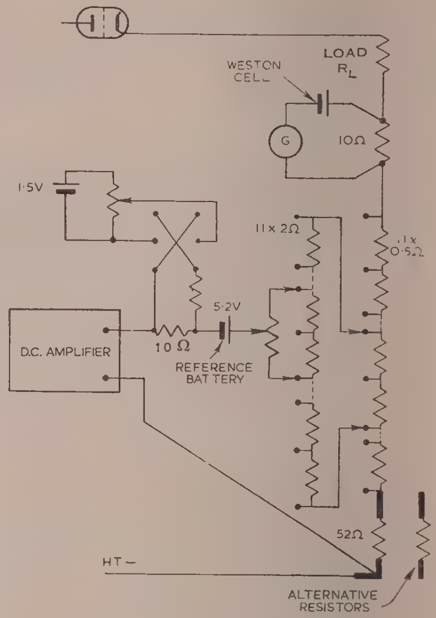


Fig. 4.—Current control circuit.

(6) TEST PROCEDURE AND RESULTS

Typical test procedure and the results obtained are as follows:

(*a*) With the main current supply off but amplifier supplies on and reference battery replaced by a short-circuit, the amplifier output is observed. Variations in this output indicate varying thermo-electric potentials in the reference circuit and amplifier input. The total variations are normally less than 5 μV, corresponding to about 1 part in 10^6.

(*b*) With the stabilizer in operation, the variation of voltage at the input of the contact-modulator amplifier, as indicated by its output voltmeter, is measured. The normal value is about 0·5 μV, which is equivalent to load-current variations of 1 part in 10^7. As discussed in the companion paper describing the high-voltage stabilizer,[1] this indicates that the feedback circuit is operating satisfactorily with adequate loop gain, and that any residual variations greater than 1 part in 10^7 must be due to variations in the reference resistor or battery.

(*c*) A 10-ohm resistor is connected in series with the load, as shown in Fig. 4, and the voltage drop across it compared with that from a miniature Weston cell by a sensitive galvanometer. The resistor is of similar design to the reference resistor and has potential terminals. The resistor and cell are located in a constant-temperature bath, so that instability due to temperature changes should amount to less than 1 part in 10^6. The galvanometer deflection is recorded by a simple photocell-amplifier arrangement[18] and pen recorder. The sensitivity of the indication is calibrated by altering the reference voltage as described in Section 5. If the checking resistor and cell are perfectly stable, this deflection indicates fluctuations in the output current. If not, the output current variation is smaller than that indicated. The ripple voltage across the load is monitored with a cathode-ray oscillograph. The impedance of the electron-microscope objective lens is 12 000 ohms at 50 c/s, so that a ripple of 1 part in 10^6 corresponds to 1·25 mV.

The effects of input-voltage and load-resistance changes were measured and found to be 0·04 part in 10^6 per 1% input-voltage change, and 0·3 part in 10^6 per 1% load-resistance change. These agree approximately with eqns. (2) and (3). The ripple voltage across the lens is 1·5 mV, i.e. 1·2 parts in 10^6 of the load current. The stability of the load current for a constant load resistance, measured as described in paragraph (*c*), is shown in Fig. 5, a calibration voltage of 2 parts in 10^6 being

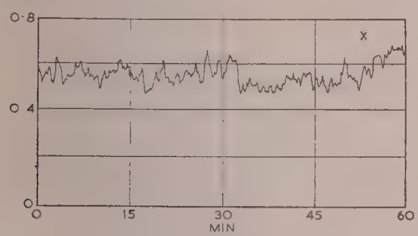


Fig. 5.—Typical chart record of output current stability.

A calibration step of 2 parts in 10^6 was applied at X; 0·2 on the chart corresponds to 4 parts in 10^6.

introduced at X. It can be seen that the peak-to-peak fluctuation over one hour is about 4 parts in 10^6.

(7) ACKNOWLEDGMENTS

The author is indebted to Mr. M. E. Haine for suggestions and helpful discussion during the development of the stabilizer, and to Dr. T. E. Allibone, F.R.S., Director of the A.E.I. Research Laboratory, for permission to publish the paper.

(8) REFERENCES

(1) HAINE, M. E., and JERVIS, M. W.: "The Use of an Electron Velocity Analyser to stabilize a 50-kV Direct-Voltage Source to a few parts in a million" (see page 265).
(2) WYNN-WILLIAMS, C. E. "An Automatic Magnetic Field Stabilizer of High Sensitivity," *Proceedings of the Royal Society, London*, A, 1934, 145, p. 250.
(3) THOMAS, H. A.: "Nuclear Resonance Magnetic Field Control," *Electronics*, 1952, **25**, p. 114.
(4) THOMAS, H. A., DRISCOLL, R. L., and HIPPLE, J. A.: "Measurement of the Proton Moment in Absolute Units," *Physical Review*, 1950, **78**, p. 787.
(5) JERVIS, M. W.: "Direct-Current Stabilizers for Electromagnets," *Electronic Engineering*, 1954, **26**, p. 429.
(6) JERVIS, M. W.: "The Temperature Coefficient of the E.M.F. of Kalium Cells," *Journal of Scientific Instruments*, 1953, **30**, p. 60.
(7) SENETT, W. P., and PIERCE, R. W.: "Exceptionally Stable Regulated Power Supply for Electrometer Tubes," *Review of Scientific Instruments*, New York, 1952, **23**, p. 534.
(8) KANDIAH, K., and BROWN, D. E.: "High-Gain D.C. Amplifiers," *Proceedings I.E.E.*, Paper No. 1196 M, November, 1951 (**99**, Part II, p. 314).
(9) LAWSON, J. L., and TYLER, A. W.: "A Magnet Current Stabilizer," *Review of Scientific Instruments, New York*, 1939, **10**, p. 304.
(10) BLACKMAN, P. F., and SUNVIC CONTROLS LTD.: British Patent No. 657690.
(11) BODE, H. W.: "Network Analysis and Feedback Amplifier Design" (Van Nostrand, 1945).
(12) BODE, H. W.: "Relations between Attenuation and Phase in Feedback Amplifier Design," *Bell System Technical Journal*, 1940, **19**, p. 421.
(13) WILLIAMS, F. C., NOBLE, S. W., COOKE-YARBOROUGH, E. H., and WHITE, E. L.: British Patent No. 647983.
(14) BUCKERFIELD, P. S. T.: "The Parallel-T D.C. Amplifier," *Proceedings I.E.E.*, Paper No. 1361 M, October, 1952 (**99**, Part II, p. 497).
(15) SOMMERS, H. S., WEISS, P. R., and HALPERN, W.: "Magnet Current Stabilizer," *Review of Scientific Instruments, New York*, 1951, **22**, p. 612.
(16) JACOBSEN, A., and HOLDAM, J. V.: "Electronic Instruments," M.I.T. Radiation Laboratory Series, **21**, Chap. 16 (McGraw Hill, 1948).
(17) HARRIS, F. K.: "Electrical Measurements" (Chapman and Hall, 1952).
(18) JERVIS, M. W.: "A Simple Photocell Amplifier Attachment for use with Galvanometers," *Journal of Scientific Instruments*, 1954, **31**, p. 302.

DISCUSSION ON THE ABOVE TWO PAPERS BEFORE THE MEASUREMENTS SECTION, 14TH DECEMBER, 1954

Professor E. Bradshaw: Since constant magnetic field strength is the aim, and since this depends on factors other than exciting current, have the authors considered whether proton magnetic resonance could be used in the field to be controlled? Secondly, have they considered whether the electron-microscope beam itself could be used for the control of the microscope accelerating voltage, by periodic deflection of the beam over a monitoring slit?

Mr. R. D. Watts: I shall refer first to the paper by Mr. Jervis.

The author has considered the various types of amplifier which might be used for the feedback amplifier on his stabilizer, and has reduced the possibilities to two permissible alternatives—the d.c. chopper-amplifier and the galvanometer-amplifier. From my own experience of both, the galvanometer type is greatly to be preferred and I am rather surprised at the author's choice of the alternative.

The d.c. chopper-amplifier can have a very excellent short-term stability when carefully adjusted, but its stability depreciates very rapidly. I attribute this to wear of the contacts and possibly to dust between them. An added disadvantage is the ripple produced by the chopper action, which cannot be overcome by the use of feedback alone.

Mr. Jervis apparently discarded the galvanometer-amplifier on the grounds that it was fragile, sensitive to vibration and subject to damage by overload. If a dynamically balanced suspension is used, i.e. the mirror of the galvanometer is fixed at the centre of the coil which is itself symmetrically suspended, and if at the same time the suspension is immersed in a liquid of the same density as the coil system, the result is an instrument which has proved itself over very many years in this particular application. Such galvanometers have been in continuously satisfactory use under much more severe conditions of shock, vibration and tendency to overload than would be met in the stabilizer application described. They are, in fact, robust. If a galvanometer suspension fails it is no more difficult to replace it than it is to replace a valve in a d.c. chopper-amplifier. In order to monitor the stability of his apparatus the author uses a galvanometer-amplifier. If a galvanometer is suitable for this purpose, it is equally suitable as the indicator of the stabilizer.

Concerning performance, it is stated that the figures given are for short-term stability. Because of my particular interest in precision stabilizers having long-term stability, I should like to know whether figures are now available giving an indication of this.

What adjustments to the d.c. chopper-amplifier are necessary to maintain the performance, and what special precautions were taken to age the manganin coils and to minimize the temperature coefficient?

With regard to the paper by Mr. Haine and Mr. Jervis, it seems that there may be a demand for an electron-optical-controlled stabilizer if there are special virtues in this system. Are there any special advantages of this method which would make worth while the production of a special electron-optical tube of simpler construction than that designed by the authors, for precision stabilizer applications?

Mr. F. C. Widdis: I am surprised that Mr. Jervis uses a contact-modulated amplifier as the detector in his stabilizer, particularly since he is working under laboratory conditions. The contact-modulated amplifier is an elaborate and expensive device usually only applied to the detection and measurement of signals of the order of $0 \cdot 1\,\mu V$ and below, where, with careful contact design and maintenance, zero stabilities of the order of $0 \cdot 01\,\mu V$ have been achieved over periods of some hours. In this region galvanometers suffer from severe limitation. However, in d.c. stabilizers of the type described the signal level is relatively large and the d.c. galvanometer forms an attractive alternative to the contact-modulated amplifier.

For some years now I have used taut-suspension galvanometers with photocell attachments on simple anti-vibration mountings in d.c. stabilizers. I have had no trouble with vibration and I cannot recollect breaking a suspension. The heavy feedback characteristic of these stabilizers seems greatly to reduce the effects of shock. I do not feel, therefore, that the author's objection to the galvanometer-amplifier on the grounds of fragility and sensitivity to vibration is valid in this instance. The contact-modulated amplifier would, of course, be preferable if the apparatus were to be used in a moving vehicle. In the past, ripple has been the main difficulty encountered in this type of stabilizer, as this, owing to the limited frequency response, cannot be controlled by the stabilizing circuit. A variety of methods have been adopted to eliminate the ripple before it reaches the stabilizing circuit, but the parallel a.c. amplifier adopted by the author seems to be the best solution to this problem. Its introduction would be a valuable improvement in precision d.c. stabilizers which use standard cells as the reference source, and are intended to give accurately known voltages for testing purposes.

THE AUTHORS' REPLY TO THE ABOVE DISCUSSION

Messrs. M. E. Haine and **M. W. Jervis** (*in reply*): The use of proton resonance for stabilization of the magnetic field in the magnetic analyser was considered but rejected, owing to the low resolution obtainable at the low field strengths required. In the case of the objective lens, the field is also non-uniform.

The possibility of stabilizing the microscope voltage supply by measuring the energy of its own beam was also considered. In theory, it would be possible to separate part of the beam and analyse this, but in practice there are so many critical problems associated with the electron microscope that any complication of its structure is not to be recommended.

Mr. Watts and Mr. Widdis compare the galvanometer and contact-modulator-type amplifiers. When built with special liquid-immersed galvanometers and with adequate vibration-proof supports, some of the objections to the galvanometer amplifier disappear, and then there is probably little to choose between the two types.

The zero drift of the contact-modulator amplifier is $\pm 0 \cdot 25\,\mu V$ over periods of one hour, and $\pm 1\,\mu V$ over periods of a week. With the contact-modulator-coil excitation and contact spacing correctly adjusted, this performance can be maintained over periods of years; in fact, the commercially produced instrument used has not had its cover removed during two years of operation. No doubt difficulty would be encountered if a stability of $\pm 0 \cdot 01\,\mu V$ were required, as suggested by Mr. Widdis, but, as mentioned in Section 3.3, this is not necessary and a stability of $\pm 1\,\mu V$ corresponds to 1 part in 5×10^6, which is quite adequate. We see no reason why these amplifiers should be applied only in the region of $0 \cdot 1\,\mu V$ and below. A galvanometer amplifier was used as the stability monitor because it was easier to insulate from earth. Variable leakage from G to earth (Fig. 4) would cause fluctuations in the effective reference resistance.

No long-term stability tests have been made on the complete stabilizer, short-term stability only being required. For this reason, no special steps were taken to age the manganin coils, but they were baked to relieve winding stresses. The temperature coefficient of 5 parts in 10^6 per degree centigrade was given by the manufacturers, and this was confirmed by a rough experiment.

There may be some other applications for the electron-optical type of stabilizer, particularly in the field of h.v. measurement. The simplifications possible are a fully-temperature-compensated permanent magnet and a sealed-off vacuum system.

621.396.67 : 621.396.812

Paper No. 1714 R
Sept. 1954

THE VERTICAL RADIATION PATTERNS OF MEDIUM-WAVE BROADCASTING AERIALS

By H. PAGE, M.Sc., and G. D. MONTEATH, B.Sc., Associate Members.

(*The paper was first received* 18*th February, and in revised form* 1st *June*, 1954. *It was published in September*, 1954, *and was read before the* RADIO SECTION 1*st December*, 1954.)

SUMMARY

The fading-free range of a medium-wave transmitting station is determined by interference between the ground wave and waves reflected from the ionosphere. In order to achieve as great a range as possible, it is common practice to reduce the strength of the reflected waves by using vertical aerials between 0·5 and 0·6 wavelengths high. The degree of success achieved is controlled by four factors: the current distribution on the aerial, the conductivity of the ground, the flatness of its surface, and the nature of reflection at the ionosphere. An experimental investigation into these has included measurements on small-scale models and on a mast-radiator now in service.

In the absence of an exact derivation of the current distribution, a semi-empirical solution has been used in computing theoretical radiation patterns. These are in reasonably good agreement with the experimental results, both for perfectly-conducting and imperfectly-conducting ground.

Unevenness of the surface of the ground was found to have a more serious effect than had generally been realized. A theoretical treatment of this problem has been verified experimentally in an idealized case, but further investigation is desirable. Another effect that has hitherto received insufficient attention is the diffuseness of ionospheric reflection. This may degrade the performance of an anti-fading aerial, particularly if an attempt is made to achieve a vertical radiation pattern with a sharp minimum.

LIST OF SYMBOLS

h = Height of aerial.
H = Height of hill.
Z_0 = Characteristic impedance of the aerial.
α = Radius of aerial, regarded as a cylinder.
ϵ = Complex relative permittivity of the ground.
θ_0 = Angle to the vertical at which the field strength is a minimum.
θ_s = Parameter describing the diffuseness of ionospheric reflection.
λ = Wavelength.

(1) INTRODUCTION

In view of the increasing congestion in the medium-wave band, it is becoming more and more important for broadcasting organizations to achieve the maximum coverage with each allocated wavelength. The potential service area may be limited either by interference or by selective fading. The former may be overcome by using a sufficiently powerful transmitter (where permitted by international agreement), but selective fading, which results from the simultaneous reception of a wave propagated along the ground and waves reflected from the ionosphere, is not affected by increasing the transmitter power; this would merely increase the ground wave and the reflected waves in the same ratio. The only way of extending the service area is to reduce radiation towards the ionosphere by controlling the vertical radiation pattern (v.r.p.) of the aerial, i.e. the pattern displaying the radiated field strength as a function of the angle from the vertical.

The simplest anti-fading aerial is a vertical wire or mast energized at the base. The v.r.p. of such an aerial may be calculated readily if the following simplifying assumptions are made:[1]

(*a*) The distribution of current along the aerial is a sinusoidal standing-wave pattern, corresponding to a velocity of propagation equal to that of light.
(*b*) The site is flat and perfectly conducting.

Assumption (*a*) is a reasonably good first approximation; it is equivalent to treating the aerial as one conductor of an open-circuited air-spaced transmission line. Assumption (*b*) is also approximately correct at medium wavelengths, since where possible transmitting stations are sited on flat ground of reasonably high conductivity.

The calculated v.r.p.'s of vertical aerials of different heights, based on the above assumptions, are shown in Fig. 1; there is a

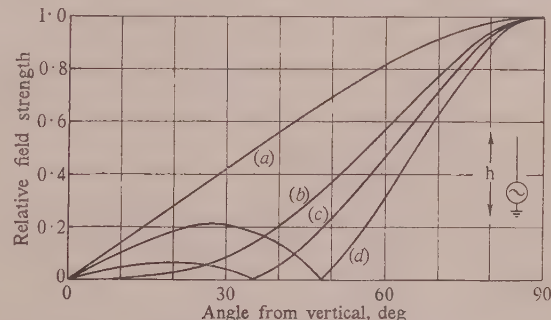


Fig. 1.—Vertical radiation patterns of vertical aerials, assuming perfectly-conducting ground and a sinusoidal standing-wave current distribution.

(*a*) $h = 0{\cdot}25\lambda$. (*b*) $h = 0{\cdot}50\lambda$. (*c*) $h = 0{\cdot}55\lambda$. (*d*) $h = 0{\cdot}60\lambda$.

zero at an angle to the vertical depending on the height of the aerial, provided that this exceeds one half-wavelength. Fig. 2 shows the field strength of the ground wave and of the reflected wave in terms of the distance from the transmitting aerial, as received by a loop aerial oriented for maximum ground-wave signal. The reflected-wave curves assume perfect reflection at a layer 110 km high. Finite ground conductivity is taken into account only in the ground-wave curve; for this a typical value of 10^{-2} mho/m is assumed, and the wavelength is taken to be 300 m. As we shall see later, a reasonable definition for the limit of the service area is that distance for which the field strength of the ground wave is equal to that of the reflected wave (calculated by assuming perfect reflection at the ionosphere). On this basis, consideration of curves similar to those shown in Fig. 2 indicates that there is an optimum aerial height giving the largest service area.[2] For the typical case considered, this optimum height is approximately $0{\cdot}56\lambda$, and the corresponding radius of the service area is 190 km—approximately twice the range achieved with a short aerial. This example is considered only to illustrate the principles; the idealized conditions assumed

Mr. Page and Mr. Monteath are with the British Broadcasting Corporation.

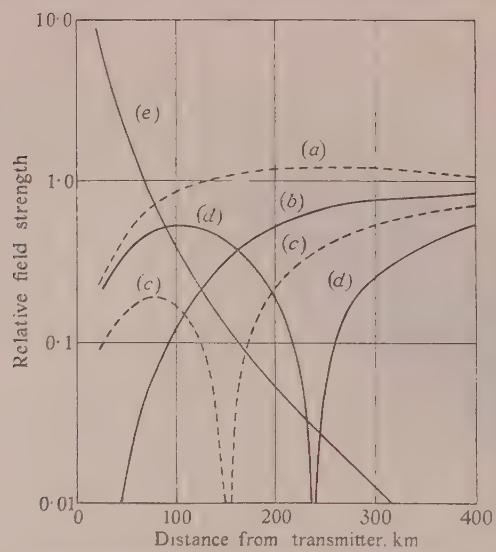


Fig. 2.—Ground-wave and reflected-wave field-strength curves corresponding to vertical radiation patterns shown in Fig. 1.

(a) $h = 0\cdot25\lambda$, (b) $h = 0\cdot50\lambda$, (c) $h = 0\cdot55\lambda$, (d) $h = 0\cdot60\lambda$: Reflected-wave field strength assuming perfectly-conducting ground, perfectly-reflecting ionosphere and idealized aerial-current distribution.
(e) Ground-wave field strength for a conductivity of 10^{-2} mho/m and a wavelength of 300 m.

do not apply in practice, and as a result the range is less than doubled by using an aerial of the optimum height.

The choice of a loop as receiving aerial in assessing the range calls for comment, since a vertical aerial would discriminate against downcoming radiation and so reduce the depth of fading. The reason is that listeners often use indoor aerials, which in respect of their sensitivity to reflected waves are no better, and may be worse, than a loop. Furthermore, the v.r.p. of a vertical aerial may be degraded appreciably by ground irregularities near the receiving point (see Section 4.3). For these reasons the loop is suggested as the most suitable type of receiving aerial to assume when estimating service areas.

The ground-wave field strength for a given power, and consequently the optimum height of the aerial, depends on the ground conductivity and the wavelength. The field strength of the ground wave at a given point is reduced if the conductivity is low or the wavelength short; the optimum height and the radius of the service area are correspondingly small. Conversely, if the conductivity is high or the wavelength long, the optimum height is greater and the radius of the service area is increased.

Typical values of optimum aerial heights over the wavelength range 200–500 m, for a ground conductivity of 10^{-2} mho/m, are $0\cdot52$–$0\cdot57\lambda$. The corresponding service area radii are approximately 80–180 km, and θ_0 (the angle of zero radiation in the curves of Fig. 1) is between 20° and 40° to the vertical.

The use of an aerial between $0\cdot5\lambda$ and $0\cdot6\lambda$ high is not the only way of increasing the radius of the service area compared with that obtained with a short transmitting aerial, but the other methods that have been proposed[3–5] either offer no advantage over a single high aerial or else involve great engineering difficulty. As a result, the only type of anti-fading aerial in practical use is a radiator between $0\cdot5\lambda$ and $0\cdot6\lambda$ high; combinations of two or more such aerials are sometimes used if a non-uniform horizontal radiation pattern is required.

The practical realization of these aerials usually takes the form of a stayed lattice mast of constant cross-section, although self-supporting masts are sometimes used. The cost of such mast-radiators is considerable, and in practice the required physical

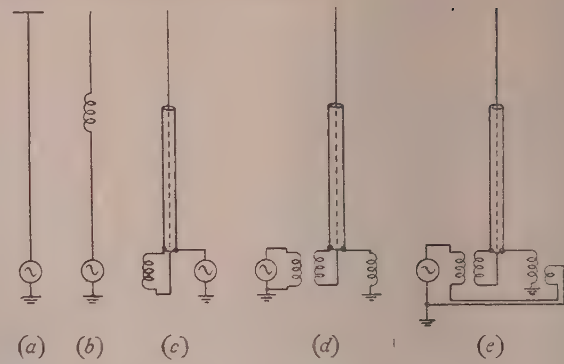


Fig. 3.—Methods of energizing vertical aerials.

(a) Aerial with capacitive top. (b), (c) Series-loaded aerials. (d) Loop-fed aerial. (e) Doubly-fed aerial.

height is reduced either by adding a capacitive top, as shown schematically in Fig. 3(a), or by breaking the mast with an insulator and loading with a series inductance as shown in Figs. 3(b) and 3(c), or by a combination of both methods. By changing the size of the capacitive top, or the value of the loading inductance, some control of the v.r.p. is obtained. This facility is useful, both in order to achieve the optimum performance and also to provide some flexibility of adjustment in the event of a change in the operating wavelength.

In the arrangements shown in Figs. 3(a), 3(b) and 3(c) the mast is energized at the base; in this condition it will be termed a "base fed" mast. Another method,[6] shown in Fig. 3(d), is to break the mast by means of an insulator at a point near to the maximum of the sinusoidal current distribution (or current loop), and to energize it at this point through a transmission line within the lower part of the mast; this arrangement will be referred to as a "loop fed" mast. The v.r.p. can be controlled by connecting an adjustable reactance between the base of the mast and earth; this is like the arrangement shown in Fig. 3(c) but with the loading point and the driving point interchanged. In yet another arrangement, shown in Fig. 3(e), the mast is energized at the break and at the base simultaneously;* this will be referred to as "doubly fed." Both the loop-fed and doubly-fed aerials have v.r.p.'s superior to that of the base-fed aerial. The way in which this improvement results from the different current distribution on the aerial is discussed in Section 2.1 for the loop-fed aerial, and in Section 3.1 for the doubly-fed aerial.

If the current distribution on the aerial were known and the earth were spherical and perfectly-conducting, the v.r.p. could be predicted accurately. If in addition the ionosphere behaved at all times like a perfect reflector, the depth of fading could be calculated. Thus, if a given depth of fading were specified as tolerable, the limits of the service area would be known precisely.

The ground-wave field strength can be measured quite accurately, but it is more difficult to calculate, or to measure with any degree of accuracy, the field strength of the reflected wave, which governs the fading. This is because the idealized conditions assumed do not obtain in practice, and the effects of deviations from them are difficult to predict. The purpose of the paper is to summarize present knowledge of the factors governing the field strength of the reflected wave, and to indicate in what directions further work appears desirable. The relevant factors, which are discussed in succeeding Sections, are

(a) The distribution of current on the aerial.
(b) The conductivity of the site, which may be modified locally by the presence of an earth system.
(c) Site irregularities, i.e. departures from flatness.
(d) The nature of ionospheric reflection.

* Vertical aerials energized at two points have been proposed by Wilmotte.[7]

Finally, consideration is given to the performance of a mast-radiator operating on a wavelength of 464m. Conclusions are drawn regarding the most suitable aerial and type of site for a medium-wave transmitting station.

(2) THE RADIATION PATTERN OF A VERTICAL AERIAL OVER FLAT PERFECTLY-CONDUCTING GROUND

In this Section a vertical aerial radiating over flat and perfectly-conducting ground will be considered. If the distribution of current along the aerial were known the v.r.p. would be known precisely. But, in fact, there is no general rigorous solution for the distribution of the current, although many workers have investigated the problem, both theoretically and experimentally.

(2.1) Previous Work on the Distribution of Current along the Aerial

The only rigorous solution for the current distribution on a transmitting aerial is that obtained by Chu and Stratton[8] for an aerial in the form of a prolate spheroid, but this idealized shape is very different from that of a practical aerial. An alternative approach is to take a practical shape and to approximate the theoretical analysis; a number of different attempts have been made along these lines.[9–15] Other workers[16–18] have carried out current-distribution measurements, but give insufficient comparisons with corresponding theoretical results.

All the theoretical treatments are in qualitative agreement on the manner in which the true current distribution differs from the sinusoidal distribution, but there is some doubt about the magnitude of the differences.

The current may be regarded as being composed of the sum of a primary current (a sinusoidal standing-wave pattern with a velocity of propagation equal to that of light, as assumed in Section 1) and a correcting term which may be called the secondary current. The secondary current may be resolved into two components, one in phase and one in quadrature with the primary current. The primary current and the in-phase component of the secondary current may together be regarded as a modified primary current, i.e. a standing-wave pattern substantially sinusoidal in shape but corresponding to a velocity of propagation less than that of light; the ratio of this velocity of propagation to that of light will be termed the "velocity factor." The velocity factor decreases, and the quadrature component of secondary current increases, as the ratio of aerial length to radius decreases.

To a great extent an aerial behaves like a transmission line terminated in a small resistance (at the current loop). The higher the characteristic impedance, the more nearly does the current distribution correspond to a pure standing wave with a velocity of propagation equal to that of light. Although the characteristic impedance of an aerial cannot be defined exactly, it appears to be sufficiently well defined for practical purposes, provided that the aerial is not too thick. Of several formulae put forward, that given below seems in best accord with experimental results. It derives from an expression due to Howe[19] for the capacitance of a vertical cylindrical aerial over perfectly-conducting ground and is

$$Z_0 = 60[\log_\varepsilon (h/\alpha) - 1] \text{ ohms}$$

where h is the length and α is the radius of the cylinder.

The effect of the reduced velocity of propagation along the aerial can be corrected by a small change in the height, since any co-phased current distribution likely to be encountered in practice will result in a v.r.p. substantially of the type shown in Fig. 1. So long as the velocity factor can be predicted or measured, therefore, the effect of the in-phase component of the secondary current on the v.r.p. can be taken into account.

Böhm (see Reference 15) called the quadrature component of secondary current the "feed current," since it is in phase with the driving voltage and therefore associated with the flow of power. The feed current is greatest near the driving point and tapers towards the top of the mast; its principal effect is to fill in the minimum between the main and minor lobes of the radiation pattern. This leads to a reduction in the service area and is particularly important when mast-radiators, rather than thin-wire transmitting aerials, are used. By breaking the mast with an insulator at the current loop, and energizing it at that point, the centre of gravity of the secondary current is elevated and its harmful effect on the v.r.p. is virtually eliminated. For this reason loop-fed mast-radiators are coming into more common use.

In order to predict the v.r.p. of an aerial it is necessary to know both the velocity factor and the feed current as precisely as possible. The theoretical analyses mentioned above are in good agreement as regards the feed current, but are less so as regards the velocity factor; in any case they do not apply to the loop-fed mast. For these reasons, and also to provide a starting point for a more general investigation of the problems referred to in Section 1, the experiments described in the next Section were undertaken. Measurements were restricted to aerials of cylindrical shape, no attempt being made to simulate lattice masts, since the current distribution is not greatly affected by the shape of the cross-section or by lattice construction. One result for a lattice mast is given in Section 6.2.

In view of the doubt about the accuracy of the available theoretical solutions, the velocity factor for the comparison "theoretical" curves was chosen empirically to give the best fit with the measured result, the same velocity factor being used for all aerials having the same value of Z_0. As regards the feed current, the theoretical value obtained by Böhm[15] was used, for the simple reason that the result is in a form which is not too laborious to calculate.

(2.2) Measured Radiation Patterns

(2.2.1) Measuring Technique.

The technique of using small-scale model aerials and carrying out tests at correspondingly higher frequencies has been employed extensively in the past decade, and this method has been used for much of the work described in the paper. A frequency of approximately 400 Mc/s was chosen because the models were then small enough for convenient handling, but not so small as to cause difficulty in construction.

It is immaterial whether the aerial under test is used for transmission or reception, the choice being decided by convenience. For the measurements described in this Section it was used for reception, and the fixed transmitting aerial was a pyramidal horn energized by a dipole connected to a 50-watt transmitter. The detector was a thermocouple. The directivity of the transmitting aerial was used to reduce reflection from neighbouring objects. In order to simulate the effect of perfectly-conducting ground a symmetrical dipole was used, with the two halves corresponding to the vertical aerial and its image in the ground. This dipole was mounted horizontally and rotated about a vertical axis. The transmitting and receiving aerials were separated by approximately 20 wavelengths, and mounted two wavelengths above ground, so that the direct and reflected waves were in phase. This arrangement ensured an adequate measuring sensitivity and also reduced errors caused by tilting of the reflected wavefront. The effect of the reflected waves on the measured v.r.p. was assessed theoretically and found to be negligible. The accuracy of measurement was considered to be within $\pm 2\%$ of the maximum field strength.

Fig. 4 shows the dipole under test. It was supported on a quarter-wave short-circuited stub, which also ensured the balance to earth of the dipole irrespective of that of the thermocouple.

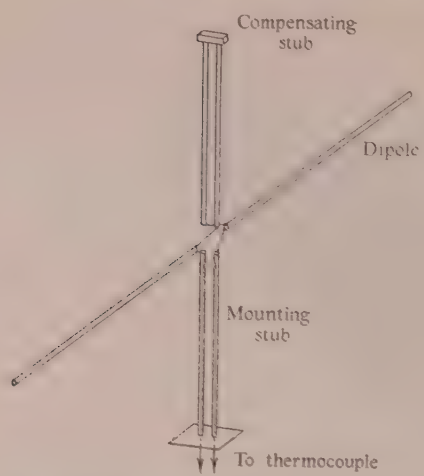


Fig. 4.—Dipole mounting for simulating measurements over perfectly-conducting ground.

Radiation from the stub was found to influence the measured v.r.p. slightly, but this effect was eliminated by fitting an identical stub above the dipole, and connecting it in the reverse direction.

The symmetry of the measured radiation pattern provided a sensitive test for unbalance between the two halves of the dipole, and for reflections from trees and other objects. The results shown are the average of the four quarters of the complete pattern; the asymmetry in a typical case did not exceed 2% of the maximum field.

(2.2.2) **Results of Measurements.**

A thin half-wave dipole was used as a standard to test the accuracy of measurement, and it was found that the measured radiation pattern agreed with the theoretical pattern to within 1% of the maximum.

Simple cylindrical aerials having characteristic impedances of 150, 245, and 460 ohms were then tested (see Section 2.1). To obtain a characteristic impedance of 460 ohms it was necessary to use a wire of 0·006 in diameter, supported by a thin wooden frame; the other models were self-supporting tubes with open ends. Measurements on each dipole were made at frequencies such that the half-lengths were $0\cdot 50\lambda$, $0\cdot 55\lambda$, and $0\cdot 60\lambda$.

The "theoretical" curves with which the measurements are compared are based on a modified primary current with an appropriate velocity factor chosen to give the best fit with the measurements; the values used were 0·95 for $Z_0 = 460$ ohms, 0·92 for $Z_0 = 245$ ohms, and 0·86 for $Z_0 = 150$ ohms. The modified primary-current distribution therefore resembled in shape that for a mast of greater length with a velocity factor of unity. The feed current for the longer mast was calculated by Böhm's[15] method, and the length scale of both primary- and feed-current distributions was contracted in proportion to the velocity factor. The fields due to the modified primary current and the feed current were then calculated and added in quadrature.

Theoretical and measured curves are compared in Fig. 5 for $Z_0 = 245$ ohms—a typical value for a mast-radiator. The field at the minimum tends to be less sharp in the case of the measured pattern, which suggests that the feed current given by Böhm's theory is too small. Agreement was better for $Z_0 = 460$ ohms, and somewhat worse for $Z_0 = 150$ ohms.

Fig. 6 shows measured v.r.p.'s of a 245-ohm series-loaded aerial $0\cdot 5\lambda$ high for two values of the loading reactance. The minimum is sharper than in the case of a simple aerial having the same angle of minimum radiation, and the difference increases with the loading reactance. This is to be expected, since the radiation resistance of a series-loaded aerial is lower, and the feed current, which is associated with the flow of power, should be smaller.

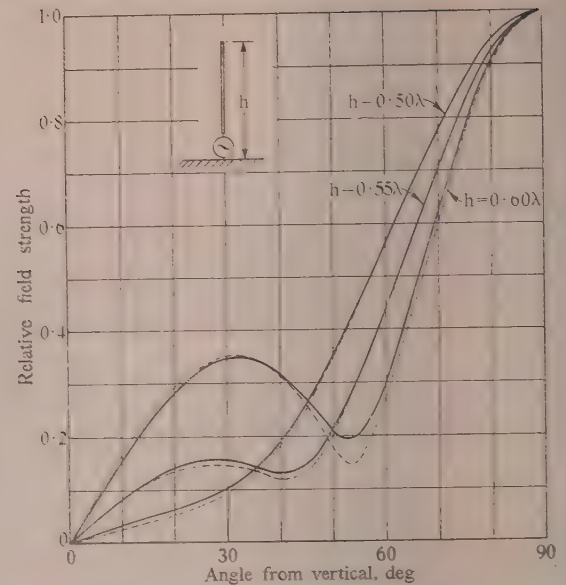


Fig. 5.—Theoretical and measured radiation patterns of base-fed aerials over perfectly-conducting ground; $Z_0 = 245$ ohms.

———— Measured.
— — — — Theoretical (velocity factor = 0·92).

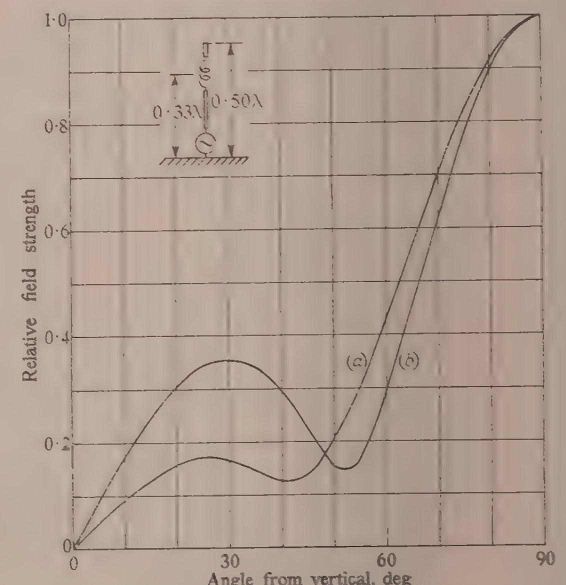


Fig. 6.—Measured radiation patterns of a $0\cdot 5\lambda$ series-loaded aerial over perfectly-conducting ground for two values of the loading reactance; $Z_0 = 245$ ohms.

(a) and (b) are for different loading reactances.

Fig. 7 shows measured v.r.p.'s of a 245-ohm loop-fed aerial for a series of values of base reactance. If Figs. 6 and 7 are compared, it will be seen that a considerable improvement in the

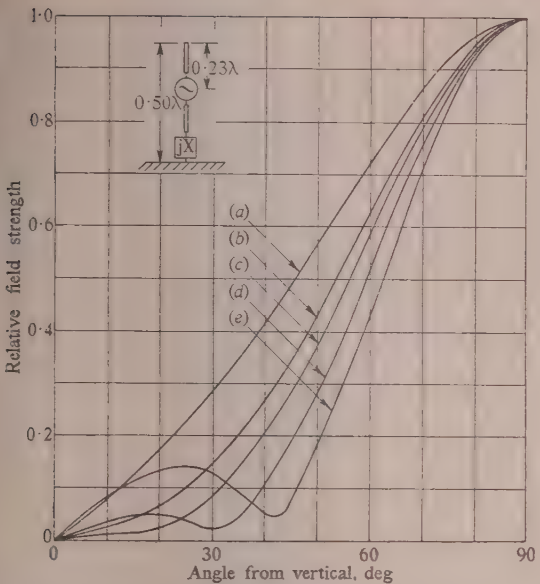


Fig. 7.—Measured radiation patterns of a loop-fed aerial over perfectly-conducting ground with different values of base reactance X; $Z_0 = 245$ ohms.

(a) $X = -133$ ohms. (b) $X = -375$ ohms. (c) $X = \infty$. (d) $X = 575$ ohms. (e) $X = 269$ ohms.

sharpness of the minimum is achieved by loop feeding. Except in the immediate vicinity of the minimum, the curves shown in Fig. 7 agree extremely well with theoretical results for a series-loaded mast carrying a co-phased sinusoidal current distribution with a velocity factor less than unity. It is possible that the results in the vicinity of the minimum are affected to some extent by limitations in the measuring accuracy.

Summarizing, the measured v.r.p.'s in typical practical cases agree reasonably well with those calculated theoretically if the velocity of propagation is chosen to give the best fit with measurements. For a cylindrical aerial having a characteristic impedance of 245 ohms (a typical value), a velocity factor of 0·92 is appropriate. The v.r.p. of a loop-fed aerial shows a considerably sharper minimum than that of the corresponding base-fed aerial.

(3) THE RADIATION PATTERN OF A VERTICAL AERIAL OVER FLAT IMPERFECTLY-CONDUCTING GROUND

(3.1) General

The term "vertical radiation pattern" requires some qualification when the ground is not perfectly conducting, since the field strength is not a product of independent functions of the distance and the direction. If the contribution to the field at any point is regarded as having two components (one associated with the space wave and one with the surface wave), the space-wave field strength is inversely proportional to the distance, whereas the surface wave is attenuated more rapidly. It follows that the resultant radiation pattern depends on the distance. The theoretical patterns compared with the measured results were therefore calculated for the particular distance used in the experiments.

In the absence of an extensive earth system, the effect of the imperfectly-conducting ground on both the space and surface waves can be calculated by well-known methods and the v.r.p. derived. The measurements without extensive earth systems, described in Section 3.3, were therefore regarded as a means of checking the measuring technique.

One of the authors[20] has described a method of calculating the effect of a local variation in the surface impedance of the ground, such as that due to an earth system. The compensation theorem for electrical networks is extended to an aerial system which includes a boundary surface exhibiting the property of surface impedance, and the change in the characteristics of the aerial system due to a small change in the surface impedance is expressed as a surface integral. Suppose, for instance, that the v.r.p. of an aerial over perfectly-conducting ground is known, and it is required to know it over imperfectly-conducting ground in the presence of an extensive earth system. The surface impedance has changed from zero to a small finite value at all points beyond the boundary of the earth system (regarding this for simplicity as equivalent to a continuous metal sheet). The change is found by evaluating the appropriate integral over this area. An application of this method indicated that practical earth systems, which do not generally exceed $0{\cdot}5\lambda$ in radius, have a negligible effect on the v.r.p. if the conductivity of the ground is at least 10^{-2} mho/m, but it was desirable to check this conclusion experimentally.

The effect of imperfectly-conducting ground on the v.r.p. is to "fill in" the zero in the same way as does the feed current in the case of a base-fed mast. In either case the additional field component responsible is in phase-quadrature with the field on either side of the minimum. Unfortunately the phase-quadrature components associated with the feed current, and with imperfect conductivity, have the same sign and so reinforce one another. But it is still possible to obtain a sharp minimum by reversing the feed current; to do this energy must be made to flow downward on the mast rather than upwards. More than the total power to be radiated is injected at the loop, and the excess is recovered at the base. For example, calling the feed current associated with a base-fed mast unity, about $-\frac{1}{2}$ unit of feed current would be required to compensate for a finite conductivity of 10^{-2} mho/m. For every kilowatt radiated, $1\frac{1}{2}$ kW would be supplied at the loop and $\frac{1}{2}$ kW would be drawn off at the base. For lower conductivities the magnitude of the reversed feed-current required would be greater.

(3.2) Measuring Technique

In order to obtain exact correspondence between the small-scale model and a full-scale aerial, it is necessary for the complex relative permittivity ϵ to be the same at both frequencies. A typical value of conductivity for medium-wave transmitting sites in Great Britain is 10^{-2} mho/m, and the corresponding value of ϵ at 1 Mc/s is $20 - j180$, in which the imaginary part is dominant.

Brine offered a convenient substance for giving the required imaginary part of ϵ, which could be controlled by adjusting the concentration of the salt; a value of ϵ of $80 - j180$ was obtained with a $3{\cdot}5\%$ solution. Some measurements were also made with a much weaker concentration (see Section 3.6). The difference between the theoretical radiation patterns for the required value of $20 - j180$, and that for the actual value of $80 - j180$ is small. Nevertheless, for comparison with the measured results the theoretical v.r.p.'s were calculated for $\epsilon = 80 - j180$.

The characteristic impedance of all aerials was 245 ohms. The distance from the base of the aerial under test was made approximately 10λ; in this case the surface wave makes a negligible contribution to the total field at angles to the vertical less than 70°. Excluding the range of angles between 70° and 90° to the vertical, therefore, the measured patterns apply at all greater distances.

Fig. 8 shows the general arrangement. A model of the aerial under test was placed at ground level and energized by a signal generator. A receiving aerial, supported by a wooden frame

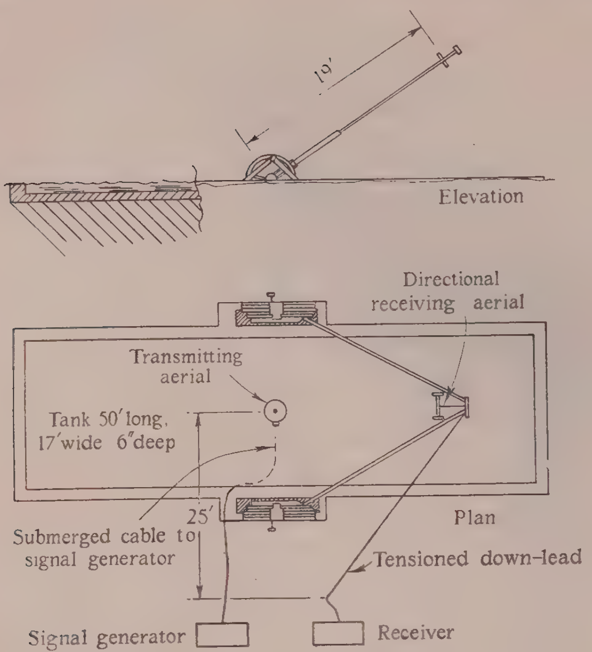


Fig. 8.—Equipment for measuring radiation patterns over imperfectly-conducting ground.

consisting of two hollow plywood poles forming an inverted "V," was arranged to rotate around this aerial at a constant radius in the vertical plane. The frame straddled a concrete tank 50ft long, 17ft wide and 6in deep, filled with brine. The aerial under test was placed in the centre of the tank and connected to the signal generator by a submerged coaxial cable, with the receiving aerial connected to a receiver by a second cable. The tank was filled to the brim, and a broadside directional receiving aerial was used; these two precautions helped to minimize reflections from the sides of the tank.

Reflection at the ends of the tank was found to cause errors only at angles within 20° of the horizontal, and separate measurements were therefore made in this region. The transmitting aerial was placed near one end of the tank, and the receiving aerial was moved vertically at the centre. Reflection from one end still contributed to the received signal, but since it was small and nearly constant in amplitude and phase, the effect on the measured v.r.p. was negligible. Corrections were made for the increase in the distance between the aerials as the receiving aerial was raised, and for the inclination of the receiving aerial to the wavefront.

The measuring procedure was to set the piston attenuator of the signal generator for a predetermined output from the receiver, so that it was unnecessary to know the law of the detector. The overall accuracy of measurement is believed to have been within $\pm 4\%$ of the maximum field.

(3.3) Base-fed Aerials without extensive Earth Systems

In order to allow comparison between experimental and theoretical v.r.p.'s, the first measurements were made in the absence of an extensive earth system. Some kind of earth connection to a grounded aerial is essential, but, provided that it is confined to the immediate vicinity of the base, it has little effect on the v.r.p. The current distribution upon which the theoretical v.r.p.'s were based was the same as that used to derive the theoretical radiation patterns shown in Fig. 5. In view of the good agreement between the theoretical and measured results for perfectly-conducting ground, the measured results for imperfectly-conducting ground should agree correspondingly well with theoretical predictions. The measurements without extensive earth systems were therefore regarded as a means of checking the measuring technique. The theoretical curves were so normalized that, had the ground been a perfect conductor, the field at the surface would have been unity. The experimental v.r.p.'s for imperfectly-conducting ground were then scaled for the best fit to the theoretical curves.

Radiation patterns were measured for aerial heights of $0\cdot25\lambda$, $0\cdot50\lambda$, $0\cdot55\lambda$, $0\cdot60\lambda$ respectively, with $\epsilon = 80 - j180$. A

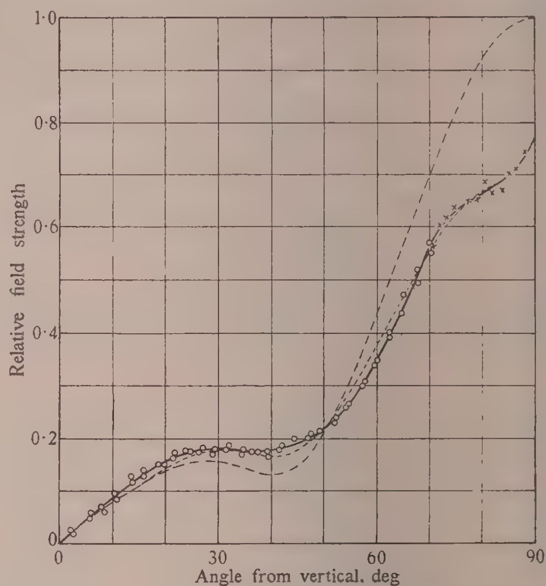


Fig. 9.—Theoretical and measured radiation patterns of a $0\cdot55\lambda$ base-fed aerial; $Z_0 = 245$ ohms.

———	Measured, imperfectly-conducting ground.
- - - - -	Theoretical, imperfectly-conducting ground.
— — —	Measured, perfectly-conducting ground.
○ ○ ○	Points measured by the method shown in Fig. 8.
× × ×	Points measured by the method described in Section 3.2.

typical result, that for an aerial height of $0\cdot55\lambda$, is shown in Fig. 9. Three v.r.p.'s are compared:

(*a*) Measured over imperfectly-conducting ground.
(*b*) Theoretical for imperfectly-conducting ground.
(*c*) Measured over perfectly-conducting ground.

The four v.r.p.'s measured over imperfectly-conducting ground are compared in Fig. 10.

(3.4) Base-fed Aerials with extensive Earth Systems

Measurements were made using earth systems consisting of perfectly-conducting circular metal sheets having radii up to one wavelength. These were level with the liquid surface.

The effect of earth systems up to $0\cdot5\lambda$ in radius was small. When the radius was increased to $1\cdot0\lambda$ there was a significant change in the v.r.p., but this was neither for the better nor for the worse, since a similar result could have been effected by a small reduction in the height of the aerial. It is of interest to note that one effect of imperfectly-conducting ground on the v.r.p. is equivalent to an increase in the height of the aerial. Correspondingly, improving the conductivity in the vicinity of the aerial by means of an earth system reduces the effective height.

The effect of asymmetric earth systems was examined by using

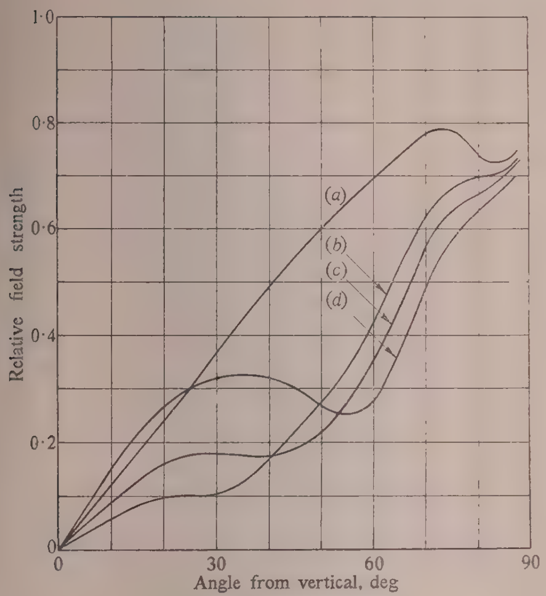


Fig. 10.—Measured radiation patterns of base-fed aerials of various heights above imperfectly-conducting ground. $Z_0 = 245$ ohms; conductivity $= 10^{-2}$ mho/m.

(*a*) $h = 0\cdot25\lambda$. (*b*) $h = 0\cdot50\lambda$. (*c*) $h = 0\cdot55\lambda$. (*d*) $h = 0\cdot60\lambda$.

a semicircular plate of radius $0\cdot5\lambda$. The field was observed near to the surface of the ground, while the earth system was rotated so as to obtain the horizontal radiation pattern. The vertical radiation pattern was then examined in the plane bisecting the semicircle. In neither case was the asymmetry of the earth system associated with a significant asymmetry in the field.

These results on base-fed aerials are consistent with theoretical predictions based on the compensation-theorem method.[20] The only important effect of the earth system is to decrease ground losses in the immediate vicinity of the aerial, thus increasing ground-wave and reflected-wave field strengths in the same ratio.

(3.5) Loop-fed Aerials

Measurements on loop-fed aerials were made by feeding energy to a break in a mast $0\cdot5\lambda$ high by a coaxial cable within the lower section. The outer conductor of this line also acted as the inner conductor of a short-circuited transmission-line stub, which provided a variable reactance between the base and earth. By moving the short-circuit v.r.p.'s corresponding to a range of effective heights were obtained. The measured v.r.p.'s, which are shown in Fig. 11, confirmed expectations in showing only small differences from the corresponding curves in Fig. 7 in the directions corresponding to the minor lobe and the minimum.

(3.6) Doubly-fed Aerials

A $0\cdot5\lambda$ aerial was fed at the base and the loop simultaneously in order to check that the predicted control of the v.r.p. could be achieved. Power was abstracted at the base by absorption in a thermistor in parallel with the variable reactance; the resistance of the thermistor was controlled remotely by passing direct current through it. The resistance and reactance were adjusted for as sharp a minimum as possible at approximately 40° to the vertical. Fig. 12 shows the resulting v.r.p., which is compared with the measured v.r.p. of a base-fed aerial of approximately the same effective height, taken from Fig. 9. A considerable improvement is obtained by substituting loop feeding

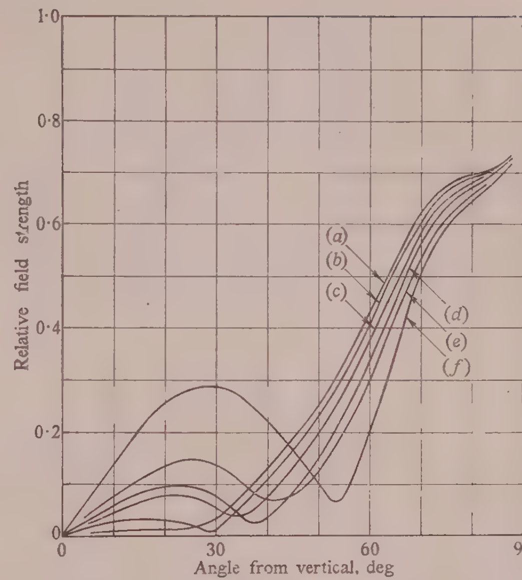


Fig. 11.—Measured radiation patterns of a $0\cdot5\lambda$ loop-fed aerial over imperfectly-conducting ground with various values of base reactance. $Z_0 = 245$ ohms; conductivity $= 10^{-2}$ mho/m.

(*a*)–(*f*) are for various base reactances.

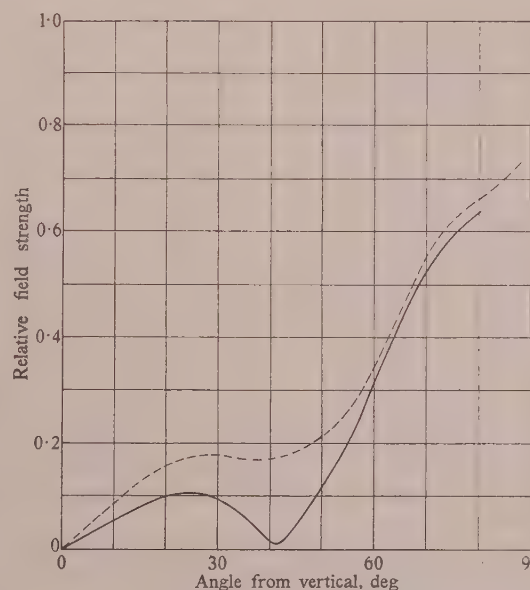


Fig. 12.—Measured radiation patterns of a base-fed aerial, and of an aerial fed simultaneously at the base and the loop; $Z_0 = 245$ ohms.

– – – – $0\cdot55\lambda$ base-fed aerial.
——— $0\cdot50\lambda$ aerial fed simultaneously at the base and the loop.

for base feeding. The additional improvement resulting from the use of combined base and loop feeding is much smaller.

When the ground conductivity was reduced, the v.r.p. of a loop-fed aerial was degraded, and as expected the improvement resulting from double feeding was more marked. A zero in the v.r.p. could be obtained at any required angle to the vertical by appropriate adjustment of the reactance and resistance connected between the base of the aerial and earth.

The conclusions to be drawn from this series of measurements are that for ground conductivity of 10^{-2} mho/m the reflected-wave field strength is substantially the same as for perfectly-conducting ground and is not affected by earth systems of practicable size. The principal effect of the finite ground conductivity is to increase the fading by attenuation of the ground wave. The v.r.p. is improved by energizing at the loop, and a further small improvement can be obtained by energizing at the base and loop simultaneously. Double feeding would become more advantageous where the conductivity is appreciably lower than 10^{-2} mho/m, but for the fact that sites of low conductivity are usually far from flat. In these cases the effect of irregularities may obscure the improvement resulting from double feeding.

(4) SITE IRREGULARITIES

(4.1) General

So far only aerials radiating over flat uniform sites have been considered. In practice sites are neither flat nor uniform, and it is important to know to what extent departures from the idealized conditions affect the radiation characteristics.

In Section 3 it was shown that the v.r.p. of a vertical aerial is not appreciably affected by imperfectly-conducting ground if the conductivity is at least 10^{-2} mho/m. Non-uniformity of the ground conductivity in different directions and at different ranges from the transmitter is therefore unlikely to be important in practical cases, in so far as the strength of the reflected wave is concerned. This is particularly true when, following the usual practice, transmitting aerials are built on sites of relatively high conductivity.

The fact that the earth is not flat but approximately spherical will also have a negligible effect on the v.r.p. at appreciable angles to the horizontal. Irregularity of the ground in the neighbourhood of the aerial is potentially much more important. For instance, an aerial mounted at the summit of a conical hill must behave, if the hill is sufficiently steep, like an aerial greater in height by the height of the hill, energized at a point corresponding to the summit of the hill.

Metzler[6] has described the results of v.r.p. measurements using a field-strength recording set carried in an aircraft. Rapid variations in the field strength with vertical angle were ascribed to waves scattered from ground irregularities; Gerber[21] also mentions this effect, but neither of these workers appear to have published a theoretical analysis of the problem; Gerber did, however, estimate the effect of re-radiation from forests. Although it is possible to postulate conditions in which extensive dense forests may affect the radiation pattern significantly, such conditions do not obtain in Great Britain. It is thought that, in this country, the only important site irregularity is the unevenness of the ground.

The question is: to what extent and up to what distance from the aerial are typical ground irregularities important? This aspect of medium-wave aerial design appears so far to have received little attention.

The approach outlined below was originally developed from an expression published previously [Reference (20), eqn. (4)]. In this paper the results will be justified by an alternative argument, which seems to give a more satisfying physical picture of the effect under consideration.

(4.2) Theoretical Considerations

If the site is perfectly conducting, the effect of an irregularity such as the hill shown in Fig. 13(a) may be regarded as an additional contribution to the field strength to be added to that due to the aerial on a flat site. The subsidiary radiator giving rise to this contribution is shown in Fig. 13(b); it is formed by the surface of the irregularity, carrying the same current as in Fig. 13(a), together with its image in the ground.

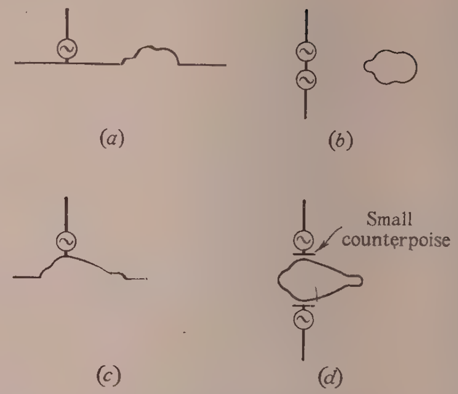


Fig. 13.—Method of calculating the effect of site irregularities on the vertical radiation pattern.

(a) Aerial on site with nearby hill.
(b) Method of calculating v.r.p. of (a).
(c) Aerial upon hill.
(d) Method of calculating v.r.p. of (c).

The difficulty is that the surface current in the presence of the irregularity cannot in general be calculated. Nevertheless, if the height of the irregularity is sufficiently small, a reasonable approximation to its effect may be obtained by assuming that the surface current is the same as for a flat site; the radiation from the subsidiary source can then be calculated. The error resulting from this procedure will be a second-order small quantity. This means that, if the vertical scale of the irregularity were reduced to zero in proportion to a scale factor x, the error in the scattered field would tend to zero like x^2.

If the aerials stand upon a hill, as in Fig. 13(c), it is more convenient to postulate the arrangement shown in Fig. 13(d).

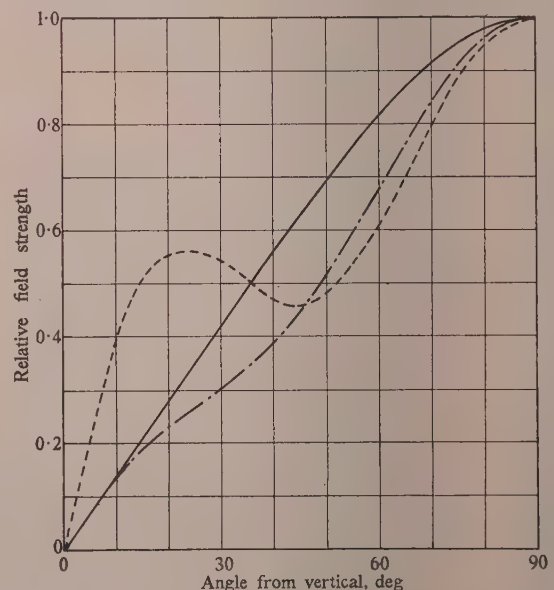


Fig. 14.—Theoretical vertical radiation patterns of a $0{\cdot}25\lambda$ aerial on a plateau and on a conical hill; perfectly-conducting site.

——— Flat site.
— — — Circular plateau, radius λ, height $0{\cdot}1\lambda$.
—·— Conical hill, radius λ, height $0{\cdot}1\lambda$.

The principal contribution to the field strength is that due to the elevated aerial in the absence of the hill, together with its image in the ground plane. It is supposed that the place of the usual earth system is taken by a non-radiating counterpoise. The contribution from the surface of the hill is calculated in the same way as for the previous case.

For irregularities exhibiting circular symmetry with respect to the transmitting aerial, the radiation is always plane-polarized and the calculations are straightforward. The analysis is summarized in the Appendix for two special cases: a $0{\cdot}25\lambda$ aerial on a circular plateau and on a conical hill. In each case the result involves an integral which must be evaluated numerically. The theoretical v.r.p.'s for these two cases, the radius being λ and the height $0{\cdot}1\lambda$ in each case, are shown in Fig. 14; they illustrate the important effect such site irregularities may have. Of interest, too, is the marked difference between the effect of the abrupt irregularity represented by the plateau, and that of the gradual irregularity represented by the conical hill.

In Fig. 15(*a*) the plateau is considered alone, for a range of heights. In order to facilitate comparison with experimental results a finite ground conductivity corresponding to a relative permittivity of $80 - j180$ was taken into account when calculating the field radiated directly by the elevated aerial. [This modified the first term of eqn. (2) in the Appendix.] Finite conductivity was, however, ignored in calculating the field re-radiated by the plateau; thus the effects of the plateau and of the finite conductivity were assumed to be additive. This procedure should be valid for small heights and directions not too near to the horizontal.

The effect of an irregularity on the v.r.p. of an anti-fading aerial may be calculated in a similar manner, but this case is more laborious to deal with because of the more complicated expression for the surface current; the surface current also depends on the method of feeding the aerial. Since an anti-fading aerial is designed to reduce radiation over a range of vertical angles, the effect of a given site irregularity on the v.r.p. will be more important than for a short aerial.

For the more complicated practical cases of aerials erected on undulating ground it is necessary to carry out a double integration numerically; the labour may be reduced by quantizing the irregularities. In general, the radiation is elliptically polarized. Some work has been done on these lines, but it is as yet too early to comment on the success of the method. Nevertheless, it is of interest to note that it seems necessary to consider the effect of irregularities at distances up to at least 5λ from the aerial, i.e. over an area of at least $75\lambda^2$.

(4.3) Measured Vertical Radiation Patterns of Aerials upon Hills

Measurements were carried out using the tank described in Section 3.2, but owing to its limited size it was not possible to make accurate measurements on the effect of irregularities extending more than about one wavelength from the aerial. Tests were restricted to surfaces which were symmetrical with respect to the aerial, namely circular plateaux and conical hills with flat tops. The models, which were made of brass, were immersed in brine having a value of ϵ of $80 - j180$; but the effect of the difference in conductivity between the plateau (or hill) and the flat part of the site is negligible compared with the effect of the irregularity itself. (This was known from the experiments with earth systems—see Section 3.4.)

Two values of plateau radius were used, namely $0{\cdot}75\lambda$ and λ; the height was varied between 0 and $0{\cdot}2\lambda$. The measured results, normalized for equal ground-wave field strength, are shown in Fig. 15(*b*) for an aerial $0{\cdot}25\lambda$ high and in Figs. 16 and 17 for aerials $0{\cdot}55\lambda$ high.

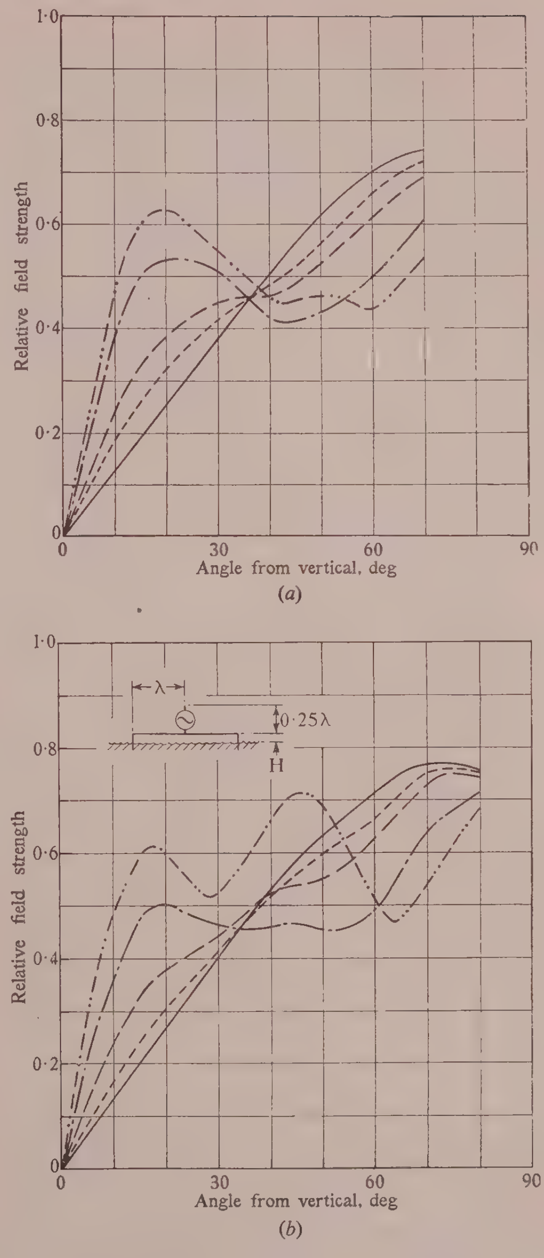


Fig. 15.—Vertical radiation patterns of a $0{\cdot}25\lambda$ aerial on a circular plateau of radius λ; conductivity of site $= 10^{-2}$ mho/m.

(*a*) Theoretical. (*b*) Measured.

———— $H = 0$.
- - - - - - $H = 0{\cdot}02\lambda$.
— · — · — $H = 0{\cdot}04\lambda$.
— · · — $H = 0{\cdot}10\lambda$.
—— —— · · $H = 0{\cdot}20\lambda$.

Corresponding measurements were made on a conical hill having a radius λ, a slope 15°, and a flat top of radius $0{\cdot}25\lambda$. Because of the shape the lateral extent of the hill was necessarily changed when the height was changed. The results are shown in Figs. 18 and 19 for aerials $0{\cdot}25\lambda$ and $0{\cdot}55\lambda$ high respectively. For either aerial the effect of a conical hill, provided its height is not too great, is similar in effect to an increase in the height of the aerial. A hill may therefore have a beneficial effect on

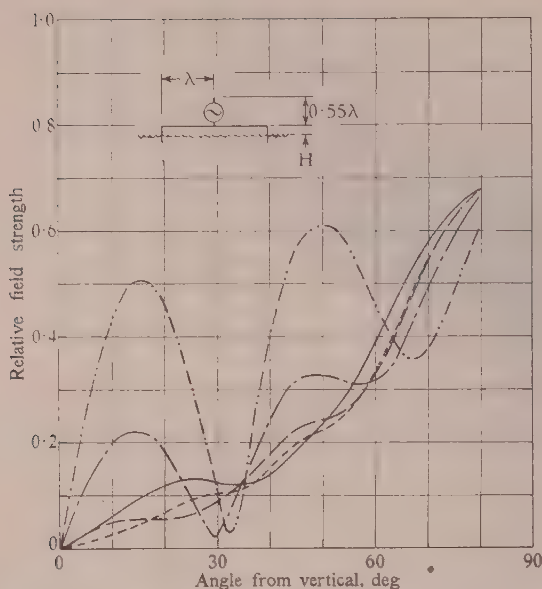


Fig. 16.—Measured radiation patterns of a $0\cdot55\lambda$ aerial on a plateau of radius λ.

$H = 0.$
$H = 0\cdot02\lambda.$
$H = 0\cdot04\lambda.$
$H = 0\cdot10\lambda.$
$H = 0\cdot20\lambda.$

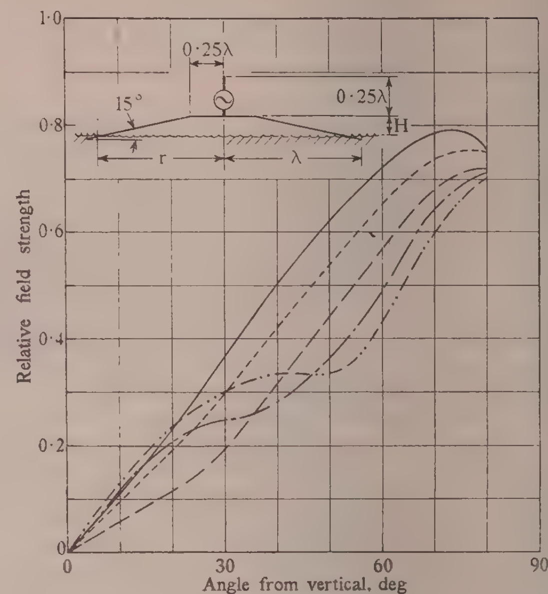


Fig. 18.—Measured radiation patterns of a $0\cdot25\lambda$ aerial upon a conical hill.

$r = 0\cdot25\lambda$, $H = 0.$
$r = 0\cdot375\lambda$, $H = 0\cdot035\lambda.$
$r = 0\cdot50\lambda$, $H = 0\cdot07\lambda.$
$r = 0\cdot75\lambda$, $H = 0\cdot14\lambda.$
$r = \lambda$, $H = 0\cdot21\lambda.$

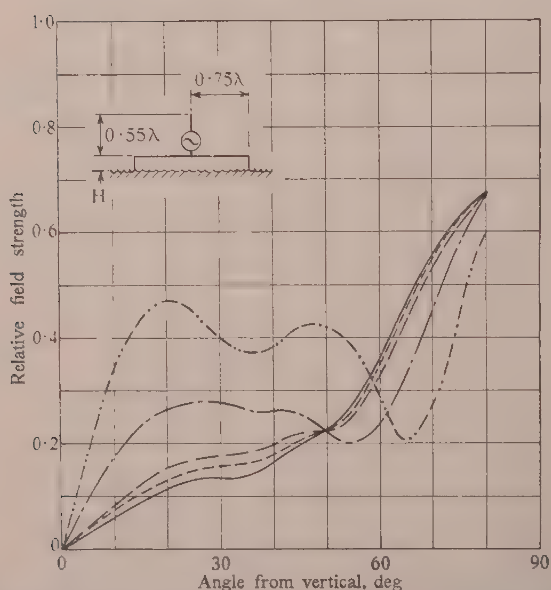


Fig. 17.—Measured radiation patterns of a $0\cdot55\lambda$ aerial on a plateau of radius $0\cdot75\lambda$.

$H = 0.$
$H = 0\cdot02\lambda.$
$H = 0\cdot04\lambda.$
$H = 0\cdot10\lambda.$
$H = 0\cdot20\lambda.$

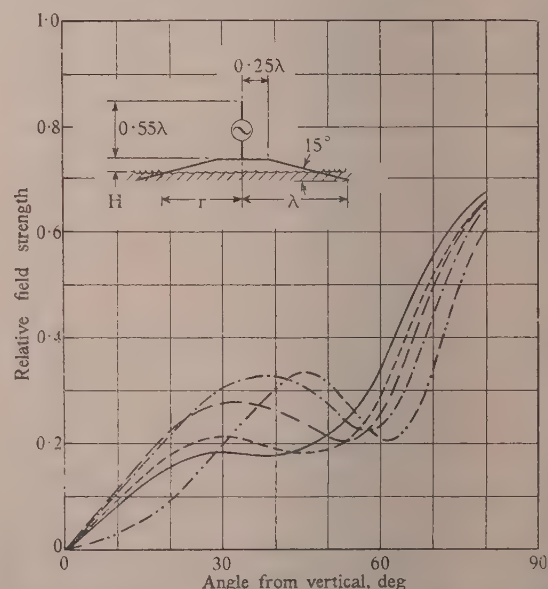


Fig. 19.—Measured radiation patterns of a $0\cdot55\lambda$ aerial upon a conical hill.

$r = 0\cdot25\lambda$, $H = 0.$
$r = 0\cdot375\lambda$, $H = 0\cdot035\lambda.$
$r = 0\cdot50\lambda$, $H = 0\cdot07\lambda.$
$r = 0\cdot75\lambda$, $H = 0\cdot14\lambda.$
$r = \lambda$, $H = 0\cdot21\lambda.$

the v.r.p. of a short aerial; in any case, as predicted theoretically, a hill with sloping sides is less harmful in its effect than a plateau of the same height.

Comparing the curves of Figs. 15(*a*) and 15(*b*), it is seen that the theoretical treatment leads to results of useful accuracy for plateaux up to $0 \cdot 1\lambda$ in height. For a height of $0 \cdot 2\lambda$ the theory leads to serious errors, particularly at angles of more than 30° to the vertical. A similar comparison between experimental and theoretical radiation patterns for plateaux of the same dimensions has been made for a $0 \cdot 55\lambda$ base-fed aerial. Because of the computation difficulties referred to above, further approximations were necessary, but the degree of agreement between theory and experiment was similar to that shown in Fig. 15. The theoretical v.r.p. for a conical hill shown in Fig. 14 is not strictly comparable with any of the experimental results of Fig. 18, but it may be seen that there is at least qualitative agreement.

The disagreement between theory and experiment for plateaux $0 \cdot 2\lambda$ high is believed to be due mainly to the effect upon the surface current of the sharp discontinuity in height at the edge. (At 1 Mc/s this corresponds to a 200ft cliff.) If consideration were restricted to the gentler gradients that are encountered in the vicinity of practical sites, it is believed that the theory would be applicable to variations of height up to about $0 \cdot 2\lambda$.

One important conclusion from the results is that site irregularities may affect not only the radiation pattern of the transmitting aerial, but also (from the reciprocity principle) that of the receiving aerial. For a short aerial on a plateau of radius λ and at least $0 \cdot 1\lambda$ high, for instance, the response is substantially the same at all vertical angles which are important as far as fading is concerned, i.e. a short vertical aerial on the plateau has the same characteristics as a loop aerial on a flat site. This is one of the reasons why the loop is suggested in Section 1 as the standard receiving aerial for the estimation of service area.

Further work on this problem appears worth while. There are two lines of attack meriting attention: an extension of the theoretical method described to an irregular site, and further measurements on small-scale models. Since irregularities up to several wavelengths from the aerial may affect the radiation pattern, such measurements should preferably be carried out at a frequency of the order of 10 000 Mc/s in order to permit the use of model sites of reasonable size.

(5) DIFFUSE REFLECTION AT THE IONOSPHERE

In Section 1 the limit of the service area was estimated by considering a single-hop path from a perfectly-reflecting layer, and then assuming that the permissible reflected-wave field strength at the receiving aerial was equal to that of the ground wave. This procedure is based on the following considerations.

Aiken[22] simulated fading by combining two identical amplitude-modulated signals of different amplitudes, the weaker being delayed in time. The distortion was most severe when the two carriers were in anti-phase; under this condition Aiken determined the maximum ratio of the amplitude of the weaker signal to that of the stronger for no perceptible distortion. Provided that the time-delay exceeded 200 microsec, as would always be the case in the service area of a medium-wave station, he found the maximum ratio to be $0 \cdot 25$.

When considering the reflected-wave field strength it is sufficient to take account of single-hop reflection at the E layer, as double-hop reflection and F-layer reflection are negligible in view of the increased path length. The most important characteristic of the ionosphere determining the reflected-wave field strength is the reflection coefficient, which may be defined for this purpose as the ratio of the field strength of the vertically-polarized component of the reflected wave to that which would exist if the layer were perfectly reflecting. From a study of the meagre published information, Ross[23] came to the conclusion that the reflection coefficient at medium wavelengths is $0 \cdot 25$ (presumably this is a median value).

This means that, if the reflected-wave field strength is predicted to be equal to the ground-wave field strength, assuming perfect reflection, its ratio to the ground-wave field strength will be $0 \cdot 25$ for 50% of the total time. Since distortion occurs mainly during the troughs of the fading, which will be assumed to occupy 20% of the fading cycle, it follows that just-perceptible distortion will occur for about 10% of the time.

The quasi-maximum value of the reflection coefficient (i.e. the value exceeded for 5% of the total time) is approximately three times the median value.[24] Deep fading will therefore occur for about 5% of the time. During the troughs of deep fading, i.e. for about 1% of the time, severe distortion will be experienced.

Unfortunately, the ionosphere does not behave like a smooth reflecting surface. There are large-scale undulations observed as a varying tilt of the reflecting layer,[25] which result in slow fading, and smaller irregularities which result in fading whose period is a few seconds. The practical implication is that the ionosphere acts as an imperfect mirror, reflecting diffusely, with slow changes in mean tilt of the reflecting surface.

Information on the magnitude of these effects at medium wavelengths is meagre. That available suggests that large-scale irregularities correspond to layer tilts of only about $\pm 1°$, which would have little effect on the effective radiation pattern of the aerial. The curvature associated with them causes a focusing effect, but this would be observed mainly as a variation in the apparent reflection coefficient. We need therefore concern ourselves only with the effect of small-scale irregularities.

A short series of observations on small-scale irregularities has been carried out in the BBC Research Department. These were made on a frequency of 863 kc/s at a range of 340km; assuming a layer at a height of 110km, the mean angle to the vertical of the reflected waves was 58°. Two receiving aerials spaced in the direction of the transmitter were set up to receive only the ionospheric reflected-wave component, the ground-wave component being suppressed. The amplitude and the relative phase of the signals at the two receiving aerials were recorded. A model[26] was assumed in which the received power density at an angle θ to the mean angle of arrival was proportional to $\exp(-\theta^2/2\theta_s^2)$. In other words, the received energy was assumed to be distributed according to a Gaussian law, the parameter θ_s being a measure of the degree of diffuseness. By examining the distribution of the relative amplitude and phase of the signal at the two aerials, a value for θ_s was derived.

Results were obtained on 18 consecutive evenings, and it cannot be assumed that conditions during this period were necessarily typical. Nevertheless it is worth recording that the results are not inconsistent with the assumed model, and that the value of θ_s was found to be about 10°. A result of similar magnitude has been obtained at the Cavendish Laboratory* from experiments made at 692 kc/s. It is, of course, desirable to conduct observations for a longer period, over a range of frequencies and distances, and it is hoped to continue the investigation along these lines.

Fig. 20 shows the effective v.r.p. of a $0 \cdot 55\lambda$ aerial, assuming diffuse reflection corresponding to $\theta_s = 10°$; it will be seen that the minimum is completely filled. If this occurs in practice, double feeding is of questionable value, since the main object of this is to achieve a sharp minimum. If, however, the conductivity is lower than 10^{-2} mho/m, double feeding obviously becomes more advantageous (assuming that the low conductivity

* Private communication from B. H. Briggs.

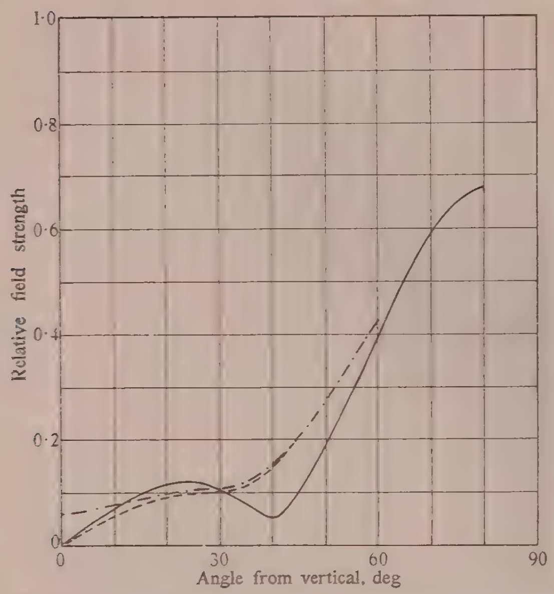


Fig. 20.—Theoretical vertical radiation patterns showing the effect of diffuse reflection at the ionosphere.

——— Vertical radiation pattern of loop-fed mast for smooth layer.
—·—·— Effective vertical radiation pattern after diffuse reflection ($\theta_s = 10°$).
------- Apparent vertical radiation pattern as measured by experimental comparison with short aerial.

is not accompanied by excessive departures from flatness). It is also clear that a small amount of scattering from the main lobe of radiation may be extremely important in determining the extent of the blurring, and in some cases this effect may make it advisable for the effective height of the aerial to be increased slightly. For this reason a study of the shape of the down-coming angular power spectrum, as well as its width, merits study.

The v.r.p.'s of anti-fading aerials are sometimes deduced by comparing the reflected-wave field strength with that radiated from a short aerial, for which the theoretical v.r.p. is assumed (for example, see Section 6.3). On the basis of specular reflection the apparent v.r.p. deduced in this way will not correspond either to the true v.r.p. (such as would be measured in an aircraft) nor to the effective v.r.p. which determines the service area. This is illustrated in Fig. 20, which shows that the apparent v.r.p., deduced in the manner described, corresponds most nearly to the effective v.r.p., differing appreciably from it only at angles to the vertical less than θ_s, a range which is not very important.

(6) MEASUREMENTS ON A MAST-RADIATOR USED FOR A BROADCASTING SERVICE

(6.1) General

This Section is concerned with the results obtained using a broadcasting aerial which was taken into service in 1951.

The requirement was to achieve as large a service area as practicable using a 150-kW transmitter on a wavelength of 464m; the transmitting equipment was to be located at the Daventry short-wave transmitting station, which was near the centre of the area to be covered. Since the large number of masts and aerials on the Daventry site would have affected the performance of a medium-wave anti-fading aerial adversely, a new site for the latter was selected, approximately 1 600m from the nearest boundary of the short-wave station. This site was as flat as could be found within a reasonable distance; the ground contours in the immediate vicinity are shown in Fig. 21. The conductivity at the site and over the greater part of the area to b covered is relatively high, being approximately 10^{-2} mho/m.

It would seem impracticable, in view of the relatively lov frequency, to use an aerial substantially higher than $0 \cdot 5\lambda$. A conventional type of mast-radiator was therefore designed, bu provision was made for double feeding. As a result of con sideration of diagrams similar to Fig. 2, the height of the mas was chosen to give a v.r.p. having a minimum at 41° to the vertical, i.e. equivalent to a radiator of height $0 \cdot 57\lambda$ having a velocity factor of unity; provision was made to obtain a range of angles of minimum radiation by adjustment of the base reactance.

The mast is of constant triangular cross-section with 9ft sides and of lattice construction. Its height is 732ft ($0 \cdot 48\lambda$) and it i broken by an insulator 470ft ($0 \cdot 31\lambda$) above ground level. Top capacitance, provided by means of six radial arms, effectively increases the height by 56ft; this value was deduced from im pedance measurements before and after erection of the top Within the lower section of the mast is a transmission line with a characteristic impedance of 100 ohms, the outer conductor being bonded to the mast at intervals and the inner to the upper section of the mast at the break. Two aerial-coupling network are provided at the base of the mast; these are arranged for eithe double feeding or base feeding, as required.

(6.2) Current-Distribution Measurements

The current distribution on the mast was measured by mean of a loop, which could be clipped to one of the vertical legs. With the mast break short-circuited, the velocity factor was found to be 0·89. This value may be compared with 0·92 deduced from the measured v.r.p.'s of small-scale models with the same value of Z_0 described in Section 2.2.2; since the models were cylin drical, it is possible that the 3% difference is associated with the lattice construction. This difference, although not large, make it worth while to check the position of the current minimum on a mast experimentally if the v.r.p. is to be known as accurately as possible. The standing-wave ratio (s.w.r.) of the measured distribution is 0·27, which is in good agreement with the value obtained by using Böhm's[15] method to compute the feed current

By loop feeding it had been expected to achieve an s.w.r. o about 0·02, but in fact it was found to be 0·09. This discrepancy is attributed to the fact that the lead-in from the inner conductor of the mast transmission line was unscreened, so that a certain amount of base feeding was taking place. It presented no difficulty in practice; a small amount of reversed base-feed was introduced to correct for the effect, and a zero s.w.r. thus obtained In this case, therefore, double feeding served the purpose o correcting imperfections in the loop-feeding circuits in an inexpensive manner; it appears a worth-while facility to retain from this point of view alone. To avoid confusion the aeria will be referred to as being loop fed in this condition.

By doubly feeding the mast, alternative current distributions could be achieved corresponding to different values of negative feed-current (see Section 3.1). But in view of the results described in Section 6.3 no experiments with negative feed current have yet been carried out.

(6.3) Measured Vertical Radiation Pattern

The v.r.p. was measured by transmitting short pulses of energy alternately from the mast-radiator and from a short aerial The ground waves from the two aerials were arranged to be equal, and the relative amplitudes of the pulses reflected from the ionosphere were measured at different distances from the transmitter. The v.r.p. of the short aerial was assumed to con form to theory for a flat site; the effective v.r.p. of the mast radiator could thus be deduced (with the limitation mentioned

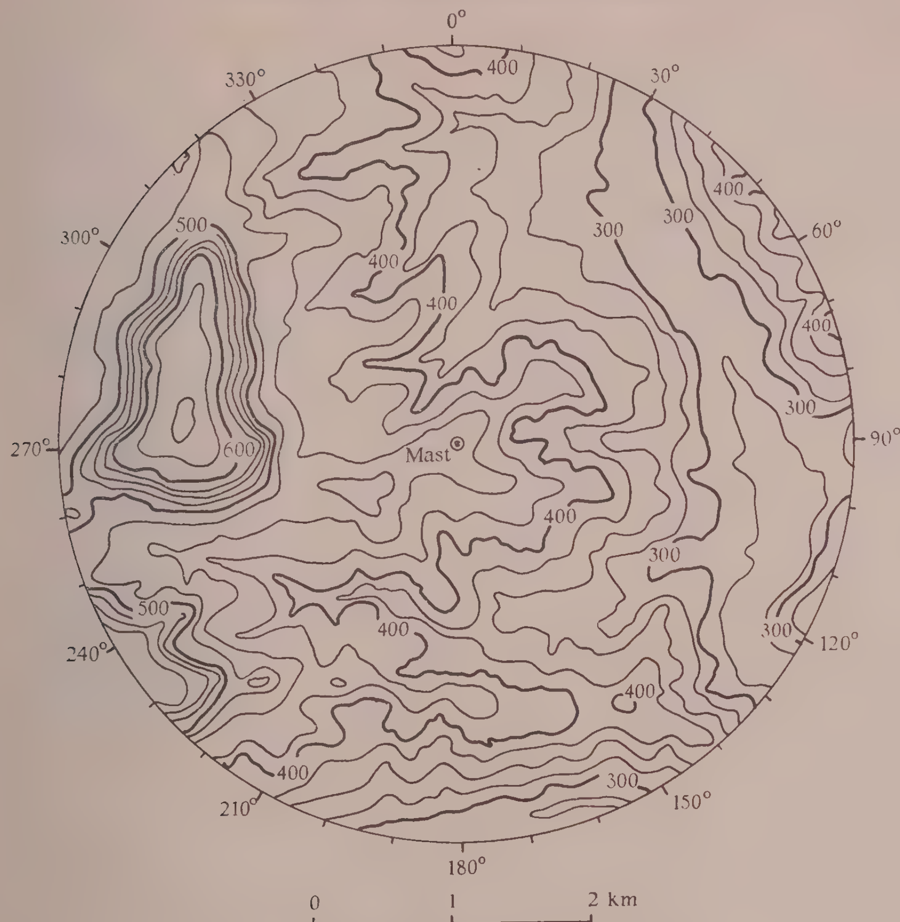


Fig. 21.—Site contours at Daventry within 6λ radius.

Contours are marked in feet.

in Section 5). Since the ground wave was always present, the delay of the echo could be measured, and the corresponding angle of incidence deduced. This comparison method has the advantage of being independent of the reflection coefficient at the ionosphere. The short aerial was erected at a distance of about 400 m from the mast-radiator and care was taken to minimize the re-radiation from one aerial while the other was pulsed.

All observations were carried out after midnight, when normal transmissions ceased. Only the strongest echo was observed; this was generally from the E layer, but sometimes from the F layer. Differential fading of the pulses received from the two aerials limited the accuracy of measurements; this diversity effect pointed to ionospheric roughness and drew attention to the importance of diffuse reflection. Under favourable conditions, i.e. reflected waves reasonably strong and differential fading not excessive, the accuracy of the results is estimated to be $\pm 30\%$. The results of measurements on the loop-fed mast in three different directions from the transmitting aerial are shown in Figs. 22–24. Individual points are indicated, the field strength being normalized with respect to the ground wave for perfectly-conducting ground. There is a significant difference between the results and the predicted v.r.p. for a flat site. Particularly is this the case on a bearing of 194° (Fig. 23), where there is a pronounced peak in the curve at an angle of about 28° to the vertical. There can be little doubt that this effect is genuine, since consistent results were obtained on three different nights; these are indicated separately in Fig. 23. Measurements on the same bearing with the mast base-fed, shown in Fig. 25, also indicate unexpectedly high radiation at about 30° to the vertical.

In view of these results it was decided to measure the relative reflected-wave field strength from the loop-fed mast over a range of horizontal directions, but at a fixed angle to the vertical which was maintained at $28\frac{1}{2} \pm 1°$. The results of these measurements are shown in Fig. 26 together with the theoretical field strength for a flat site. Individual measurements are indicated, but some of the extreme values have been avoided in drawing the curve, although it is possible that even more extreme values may have been neglected owing to the limited number of measuring points.

The relatively sharp maxima and minima in the v.r.p., and the difference in different directions from the transmitting aerial, suggest that site irregularities are influencing the performance to an important extent. As a result we would expect the reflected-wave field strength to be greater in some directions and smaller in others than the predicted value for a flat site; Fig. 26 shows this to be the case. The effect of the same site irregularities on the v.r.p. of the short aerial are not considered to be sufficiently serious to affect the deductions appreciably.

There are differences of about $0 \cdot 2\lambda$ in ground level in the

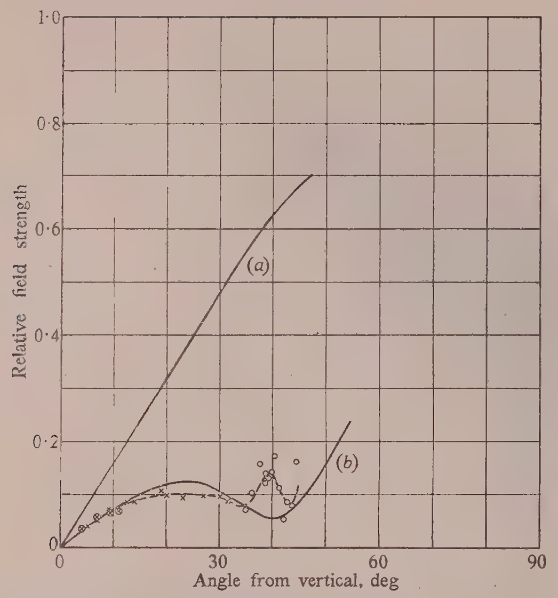


Fig. 22.—Measured vertical radiation pattern of Daventry loop-fed mast-radiator on bearing 135°.

(a) Short aerial (assumed pattern).
(b) Loop-fed mast (theoretical pattern).

Date	Single-hop E	Multiple-hop E
10/6/52	○	
11/6/52	×	⊗

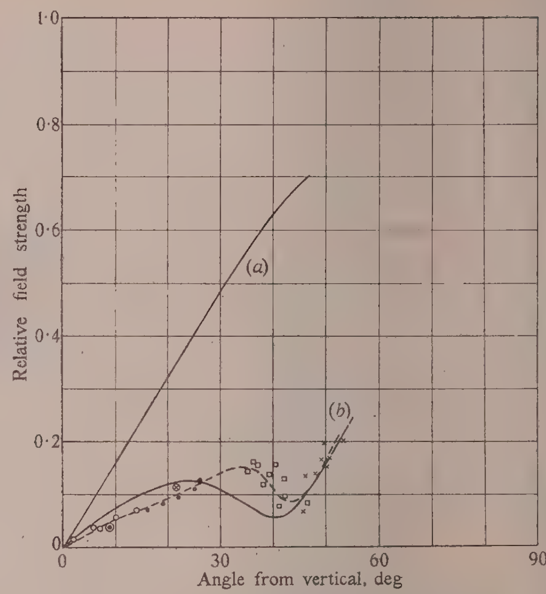


Fig. 24.—Measured vertical radiation pattern of Daventry loop-fed mast-radiator on bearing 355°.

(a) Short aerial (assumed pattern).
(b) Loop-fed mast (theoretical pattern).

Date	Single-hop E	Multiple-hop E
27/5/52	○	
28/5/52	×	⊗
29/5/52	□	
30/5/52	●	⊙

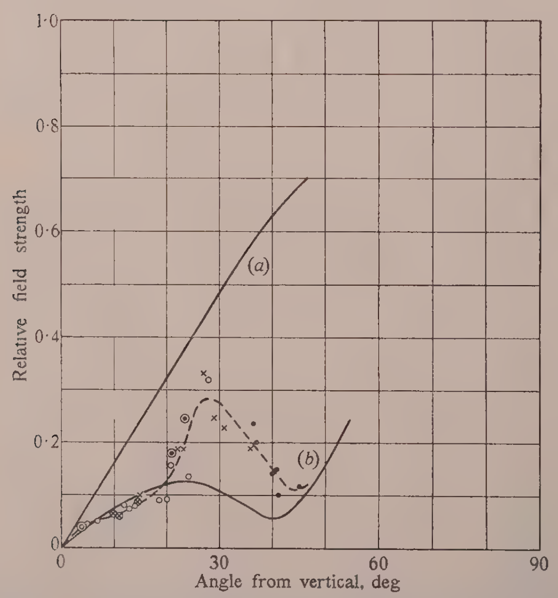


Fig. 23.—Measured vertical radiation pattern of Daventry loop-fed mast-radiator on bearing 194°.

(a) Short aerial (assumed pattern).
(b) Loop-fed mast (theoretical pattern).

Date	Single-hop E	Multiple-hop E
31/5/52	○	⊚
6/6/52	×	⊗
7/6/52	●	⊙

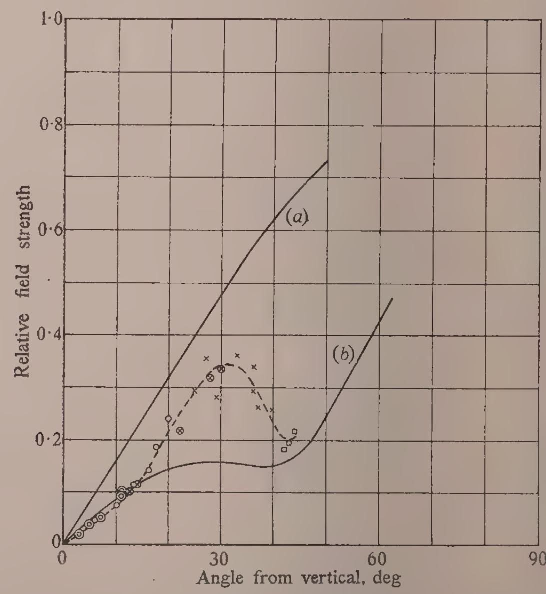


Fig. 25.—Measured vertical radiation pattern of Daventry base-fed mast-radiator on bearing 194°.

(a) Short aerial (assumed pattern).
(b) Base-fed mast (theoretical pattern).

Date	Single-hop E	Multiple-hop E
24/7/52	○	⊚
25/7/52	×	⊗
26/7/52	□	

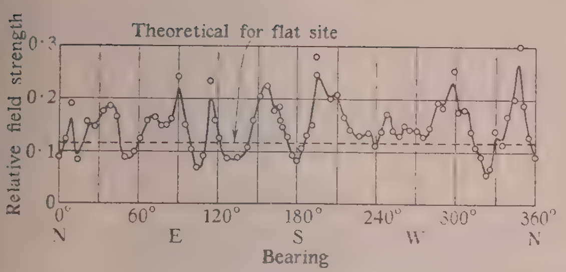


Fig. 26.—Daventry loop-fed mast-radiator; reflected-wave field strength in different directions.

—o—o—o— Measured.
— — — — Theoretical for flat site.

vicinity of the Daventry site (see Fig. 21). Now it does not seem unreasonable to require the deviations of relative field strength from the theoretical value for a flat site to be not more than one-quarter of those shown in Fig. 26. This suggests as a definition of a "flat" site for an anti-fading aerial one for which the differences in ground level do not exceed $0{\cdot}05\lambda$. The total area to be considered in this way is believed to be at least 5λ in radius. The effect of uneven ground will of course depend on the extent and abruptness of the irregularities, so that this definition should be regarded as a very rough yardstick.

Prior to the v.r.p. measurements described above, some recordings were made of service transmissions at a distance of 175 km from Daventry, on a bearing of 194°. This distance was selected as representing the fringe of the expected service area; the ground-wave field strength was 3 mV/m. The mast was loop-fed and base-fed on alternate days, and the field strength recorded during service transmissions over a period of 5 weeks (April 1–May 6, 1952) between 1800 B.S.T. and midnight. Simultaneous recordings were made using a vertical receiving aerial and a loop oriented for maximum signal; the reflected-wave component was deduced from the variation of field strength from the daytime value. Assuming specular reflection at the ionosphere, the angle of arrival could be calculated from the ratio of the reflected waves at the two receiving aerials.

The peak reflected wave during each hour was selected for analysis, the peak value being the most important from the point of view of a broadcasting service. Approximately 200 results were obtained during the recording period; the average of those obtained after 2000 G.M.T. (i.e. 75% of the total number) corresponded to a reflection coefficient of approximately $0{\cdot}5$. The ratio of the reflected wave for the loop-fed condition to that of the base-fed condition was $0{\cdot}85$, whereas the theoretical value for a flat imperfectly-conducting site (conductivity 10^{-2} mho/m) and smooth ionosphere is $0{\cdot}4$.

The angle of arrival for the strong bursts of reflected wave averaged 38° for the loop-fed condition and 42° for the base-fed condition. Although these results may be in error by a few degrees, the difference of 4° is considered significant. Individual readings fluctuated about the mean to about the same extent, $\pm 5°$ on the average. This effect, together with the different measured angle of arrival for the loop-fed and base-fed conditions, suggests diffuse ionospheric reflection. On the other hand, the shift of the main lobe of the vertical radiation pattern to the left, shown in Fig. 20 and associated with diffuse ionospheric reflection, is not revealed in Figs. 22–25. In view of the relatively few results available it is not possible to say whether this discrepancy is significant, but it seems probable that both transmitting-site irregularities and diffuse reflection at the ionosphere were influencing the performance of the loop-fed mast, with site irregularities predominating. It follows that, in this case, double feeding is unlikely materially to improve fading conditions at the fringe of the service area.

(7) CONCLUSIONS

The most common type of medium-wave anti-fading aerial in use at the present time is a mast-radiator of constant cross-section, between $0{\cdot}5$ and $0{\cdot}6\lambda$ high. If such an aerial were erected on a flat uniform site, and if the ionosphere behaved like a smooth reflecting surface, a substantially larger service area would result from loop feeding, as opposed to base feeding. A further improvement would be effected by double feeding, i.e. injecting power at the loop and drawing off part at the base, but this improvement would be small at sites of high conductivity. The lower the ground conductivity, the greater is the increase in range achieved by double feeding compared with loop feeding.

Earth systems of the usual size do not affect the v.r.p. appreciably; their main effect is to increase the efficiency by reducing ground losses. Under the assumed conditions the fading-free range of a mast-radiator could be predicted with reasonable accuracy. Conditions at the receiving site, such as the type of aerial and ground irregularities, may affect the extent of the fading but not the choice of transmitting aerial to give the optimum service area.

Two effects may degrade the performance of a mast-radiator: one is diffuse reflection at the ionosphere, which is equivalent to a blurring of the v.r.p.; the other is irregularity of the ground at the transmitting site, which causes a permanent distortion of the v.r.p. On the average an increase in the depth of fading will result, but in certain directions ground irregularity may cause a reduction in the fading. It follows that the site chosen for a transmitting aerial should be as flat and uniform as possible; the height of the undulations should preferably not exceed $0{\cdot}05\lambda$ up to a range of at least 5λ from the transmitting aerial.

Both diffuse reflection and ground irregularity reduce the effectiveness of special methods of energizing the mast, such as loop feeding and double feeding. It is, in fact, doubtful whether double feeding is ever worth-while in a practical case (except in so far as base feeding is used to correct for deficiencies of the circuits used for loop feeding) unless the ground conductivity is appreciably less than 10^{-2} mho/m.

The degree of diffuseness of ionospheric reflection and the effect of ground irregularities at the transmitting site are two aspects of medium-wave anti-fading aerial design which merit further investigation.

(8) ACKNOWLEDGMENTS

The authors wish to thank the Chief Engineer of the British Broadcasting Corporation for permission to publish the paper, and also Mr. W. Proctor Wilson, Head of the BBC Research Department, for his advice on its preparation. They also wish to acknowledge the contributions to the work made by their colleagues in the BBC Research Department. In particular Mr. R. A. Rowden was responsible for suggesting and supervising the v.r.p. measurements described in Section 6.3; Dr. G. J. Phillips was responsible for the work described in Section 5, and Mr. P. Knight was responsible for the measurements on the small-scale models, and for the theoretical v.r.p. calculations in Sections 2 and 3.

(9) REFERENCES

(1) van der Pol, B.: "On the Wavelength and Radiation of Loaded Antennae," *Proceedings of the Physical Society*, 1917, **29**, p. 269.

(2) Ballantine, S.: "High Quality Radio Broadcast Transmission and Reception," *Proceedings of the Institute of Radio Engineers*, 1934, **22**, p. 564.
(3) Jeffers, C. L.: "An Antenna for controlling the Nonfading Range of Broadcasting Stations," *ibid.*, 1948, **36**, p. 1426.
(4) British Patent 347525, April 30, 1931.
(5) Chireix, H.: "Antennes á rayonnement zénithal réduit," *L'Onde Électrique*, 1936, **15**, p. 440.
(6) Metzler, E.: "Sur un récent développement de l'antenne-pylône antifading," *Journal des Télécommunications*, 1940, **7**, p. 61.
(7) United States Patent No. 2,283,619, December 18, 1941.
(8) Chu, L. J., and Stratton, J. A.: "Forced Oscillations of a Prolate Spheroid," *Journal of Applied Physics*, 1941, **12**, p. 241.
(9) Hallén, E.: "Theoretical Investigations into the Transmitting and Receiving Qualities of Antennas," *Nova Acta* (Uppsala), 1938, Series 4, **11**, p. 1.
(10) Schelkunoff, S. A.: "Antenna Theory and Experiment," *Journal of Applied Physics*, 1944, **15**, p. 54.
(11) Gray, M. C.: "A Modification of Hallén's Solution of the Antenna Problem," *ibid.*, 1944, **15**, p. 61.
(12) Schelkunoff, S. A.: "Concerning Hallén's Integral Equation for Cylindrical Antennas," *Proceedings of the Institute of Radio Engineers*, 1945, **33**, p. 872.
(13) King, R., and Middleton, D.: "The Cylindrical Antenna; Current and Impedance," *Quarterly Journal of Applied Mathematics*, 1946, **3**, p. 302.
(14) Schelkunoff, S. A.: "Theory of Antennas of Arbitrary Size and Shape," *Proceedings of the Institute of Radio Engineers*, 1941, **29**, p. 943.
(15) Wells, N.: "Aerial Characteristics," *Journal I.E.E.*, 1942, **89**, p. 76.
(16) Morita, T.: "Current Distributions on Transmitting and Receiving Antennas," *Proceedings of the Institute of Radio Engineers*, 1950, **38**, p. 898.
(17) Barzilai, G.: "Experimental Determination of the Distribution of Current and Charge Along Cylindrical Antennas," *ibid.*, 1949, **37**, p. 825.
(18) Gihring, H. E., and Brown, G. H.: "General Considerations of Tower Antennas for Broadcast Use," *ibid.*, 1935, **23**, p. 311.
(19) Howe, G. W. O.: "On the Capacity of Radio-Telegraphic Antennae," *Electrician*, 1914, **73**, p. 829.
(20) Monteath, G. D.: "Application of the Compensation Theorem to Certain Radiation and Propagation Problems," *Proceedings I.E.E.*, Monograph No. 3, June, 1951 (**98**, Part IV, p. 23).
(21) Gerber, W., and Werthmüller, A.: "Über die Streustrahlung der Erdoberfläche im Bereich der Rundspruchsender," *Technische Mitteilungen der schweizer Telegraphen- und Telephonverwaltung*, No. 1, 1940.
(22) Aiken, C. B.: "A Study of Reception from Synchronised Broadcast Stations," *Proceedings of the Institute of Radio Engineers*, 1933, **21**, p. 1265.
(23) Ross, W.: "Ground and Ionospheric Rays," *Wireless Engineer*, 1937, **14**, p. 306.
(24) Report of C.C.I.R. Committee on Radio Wave Propagation, Bucharest, 1937; published in *Proceedings of the Institute of Radio Engineers*, 1938, **26**, p. 1193.
(25) Bramley, E. N., and Ross, W.: "Measurements of the Direction of Arrival of Short Radio Waves Reflected at the Ionosphere," *Proceedings of the Royal Society*, A, 1951, **207**, p. 251.
(26) Bramley, E. N.: "Diversity Effects in Spaced-Aerial Reception of Ionospheric Waves," *Proceedings I.E.E.*, Paper No. 1062 R, January, 1951 (**98**, Part III, p. 19).
(27) Phillips, G. J.: "The Influence of the Ionosphere on Medium-Wave Broadcasting," *The BBC Quarterly*, 1953, **8**, p. 40.
(28) Brown, G. H.: "The Phase and Magnitude of Earth Currents near Radio Transmitting Antennas," *Proceedings of the Institute of Radio Engineers*, 1935, **23**, p. 168.

(10) APPENDIX

(10.1) The Vertical Radiation Pattern of a Quarter-Wave Vertical Aerial on a Circularly-Symmetrical Hill

(10.1.1) Symbols (see Fig. 27).

All lengths are expressed in radians (2π radians = 1 wavelength). The loop current in the aerial, which provides the phase reference, is unity.

(ρ, ϕ, z) = Cylindrical polar co-ordinates of a point on the surface of the hill.
ρ_0 = Radius of the hill.
H = Height of the hill.
$r = \sqrt{(\frac{1}{4}\pi^2 + \rho^2)}$.
$r_0 = \sqrt{(\frac{1}{4}\pi^2 + \rho_0^2)}$.
A = Constant depending on the distance.
$E(\theta)$ = The electric field strength at some great distance at an angle θ to the vertical.
$I(\rho)$ = Total surface current at a radial distance ρ (positive when directed inwards).
J_0, J_1 = Bessel functions of the first kind.

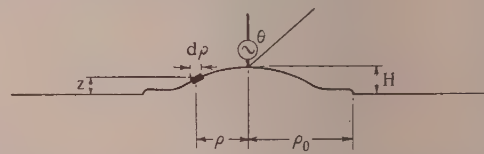


Fig. 27.—Vertical aerial on a circularly-symmetrical hill.

(10.1.2) The General Case.

The ground, including the surface of the hill, is assumed to be perfectly conducting. The first step is to determine the field strength due to radiation from the aerial above a perfectly-conducting ground-plane in the absence of the hill. For this purpose the base of the aerial is assumed to be elevated to the height of the hill, and the place of the earth connection to be taken by a small non-radiating counterpoise. To the field of the aerial will be added the field due to the current flowing in the surface of the hill, together with its image in the ground plane. The field of the surface current will be obtained as the sum of contributions from elementary annuli, and the contribution from each annulus will be resolved into parts associated with the horizontal and vertical components of the current.

At a great distance the field strength due to a vertical quarter-wave aerial with its base elevated to a height H above ground is equal to

$$A\left\{\cos\left[(H + \tfrac{1}{2}\pi)\cos\theta\right] + \cos\theta\sin(H\cos\theta)\right\}\operatorname{cosec}\theta$$

The total surface current is assumed to be the same as for a grounded aerial on a flat site. Thus[28]

$$I(\rho) = j\varepsilon^{-jr}$$

The field strength due to the vertical component of surface current in an annulus $(\rho, \rho + d\rho)$ together with its image, is

$$-jA\frac{dz}{d\rho}\varepsilon^{-jr}J_0(\rho\sin\theta)\cos(z\cos\theta)\sin\theta d\rho$$

The field strength due to the horizontal component of the surface current, together with its image, is

$$-jA\varepsilon^{-jr}J_1(\rho \sin\theta)\sin(z\cos\theta)\cos\theta d\rho$$

Summing the contributions to $E(\theta)$ from the aerial and all the elementary annuli, the result is

$$E(\theta) = A\left\{\cos\left[(H+\tfrac{1}{2}\pi)\cos\theta\right] + \cos\theta\sin(H\cos\theta)\right\}\operatorname{cosec}\theta - jA\int_0^{\rho_0}\left[\frac{dz}{d\rho}J_0(\rho\sin\theta)\cos(z\cos\theta)\sin\theta + J_1(\rho\sin\theta)\sin(z\cos\theta)\cos\theta\right]\varepsilon^{-jr}d\rho \quad . \quad (1)$$

(10.1.3) **A Circular Plateau.**

The first term of the integral is zero, except at the boundary, where z is discontinuous and there is a finite contribution to the integral. The result is

$$E(\theta) = A\left\{\cos\left[(H+\tfrac{1}{2}\pi)\cos\theta\right] + \cos\theta\sin(H\cos\theta)\right\}\operatorname{cosec}\theta + jA\sin(H\cos\theta)\tan\theta\, J_0(\rho_0\sin\theta)\varepsilon^{-jr_0} - jA\sin(H\cos\theta)\cos\theta\int_0^{\rho_0}J_1(\rho\sin\theta)\varepsilon^{-jr}d\rho \quad . \quad . \quad (2)$$

(10.1.4) **A Conical Hill.**

z decreases linearly with ρ, and $dz/d\rho$ is constant. The result is

$$E(\theta) = A\left\{\cos\left[(H+\tfrac{1}{2}\pi)\cos\theta\right] + \cos\theta\sin(H\cos\theta)\right\}\operatorname{cosec}\theta + jA\int_0^{\rho_0}\left\{\frac{H}{\rho_0}J_0(\rho\sin\theta)\cos\left[\frac{H(\rho_0-\rho)}{\rho_0}\cos\theta\right]\sin\theta - J_1(\rho\sin\theta)\sin\left[\frac{H(\rho_0-\rho)}{\rho_0}\cos\theta\right]\cos\theta\right\}\varepsilon^{-jr}d\rho \quad . \quad (3)$$

DISCUSSION BEFORE THE RADIO SECTION, 1ST DECEMBER, 1954

Mr. G. Millington: The authors have referred to the pioneer work of van der Pol on the subject of aerial polar diagrams and radiation resistance, and it is remarkable that we have had to wait nearly 40 years for a paper of the scope of the present one. I have always been impressed by the fact that the van der Pol treatment contains a paradox, namely that the energy flowing through a large surface surrounding the aerial is calculated for a reactive current distribution on the aerial.

As Böhm showed later, this difficulty can be resolved by adding the feed-current distribution to satisfy the field conditions in the neighbourhood of the aerial without modifying greatly the distant field. It is a striking fact that, if the radiation resistance is calculated from the power absorbed from the source at the aerial terminals by the combined sinusoidal and feed-current distributions, it has precisely the same value as that given by van der Pol.

Where the van der Pol method is inadequate is in describing the minima of the vertical polar diagram, and the blunting of these minima by the addition of the feed-current distribution is all-important in the consideration of the anti-fading properties of aerials. The complete theory taking into account the imperfect conductivity and the surface irregularities of the earth is prohibitively difficult, and we are indebted to the authors for an approximate treatment and for the practical investigation by the model technique.

In Section 3.1 it is stated that the radiation pattern depends on the distance, whereas the results show that it is modified in only a minor way from the pattern with a perfectly conducting earth. I should have thought that the energy which goes up to the ionosphere is mainly reflected from the earth fairly near the transmitter, and that the modification to the diagram can be largely accounted for by using the Fresnel reflection coefficient, which is a function of the angle of incidence at the ground, there no longer being a perfect image transmitter.

The ionosphere, which is so often the essential factor in propagation, is in the present case the cause of the trouble by producing an interfering wave and hence selective fading. Incidentally, the reference to this fading in the Introduction implies at first sight that it is not an interference effect, whereas the distinction that is being made is between interference from other transmitters and interference from one's own transmitter by an unwanted propagation path.

In Section 2.2.1 the reader might think that the fact that the direct and ground reflected waves are in phase is the outcome of the receiver being two wavelengths above the ground. I presume that the height of the receiver for the required phase relationship happened to be two wavelengths, but would have to be adjusted if the other parameters were altered.

At the beginning of the paper there is a reference to the use of a loop aerial for reception as being most representative of the average type of aerial used by the listener. As the authors have mentioned, the problem is reciprocal and applies equally to the design of the receiving aerial. The protection against fading would be doubled, in decibels, if anti-fading aerials were used at each end. I think that it is noble of the B.B.C. to do so much at their end when we do so little to co-operate with them at ours.

In conclusion, I should like to ask, now that all this work has been done, what its practical impact has been. Have the B.B.C. received many letters from satisfied listeners who now get improved reception? I do not ask this in order to detract from the importance of this work, because we shall all agree that in this paper we have for the first time a wealth of fundamental results that will provide the answers to many future queries.

Mr. P. P. Eckersley: I should like to refer to an early paper by myself, T. L. Eckersley, and H. L. Kirke,* not so much to add to the knowledge revealed in the paper under discussion, but because I feel that some account of the pioneering work done in those days would be of interest. This paper reveals that our intention was essentially to increase the power efficiency of a broadcasting system. It was appreciated that the more power which could be put into the ground ray, the greater would be the extent of the service area. Not much attention was paid to the vertical radiation pattern, although the point seems to have been appreciated by implication. The value of the present paper is to consider the vertical radiation pattern in detail and to combine the effects of the ground ray with it.

The previous work of Ballantine had shown that—theoretically, at any rate—the half-wave aerial was bound to be better than those commonly in use at the time. The practical work described in our paper involved experimental tests to see how far theory could be confirmed by practice. To this end an aerial was supported by a kite balloon, thereby enabling us to test any number of ratios of aerial height to wavelength.

The experiments, which were conducted about 1927, proved

* ECKERSLEY, P. P., ECKERSLEY, T. L., and KIRKE, H. L.: "The Design of Transmitting Aerials for Broadcasting Stations," *Journal I.E.E.*, 1929 67, p. 507.

without a peradventure, that the half-wave aerial had enormous advantages. The mechanical difficulties of maintaining the aerial structure sufficiently rigidly during bad weather shortened the period of tests to the extent that not a great deal of detailed knowledge could be gained from them, but the generalization that the half-wave aerial had virtue was certainly established.

In performing these experiments we naturally encountered the problem of how best to feed power into the aerial, and most of the circuits shown in the present paper were either used or adumbrated.

Some attempts were made to use Franklin's method to stack one half-wave aerial on top of another by the insertion of the wrapped artificial aerial at the join of the two open half-wave aerials. For one reason or another the scheme did not work, and as a result, the idea of feeding power into the middle of the aerial was conceived. Largely because of adverse weather conditions, we were unable to make any practical tests.

As to fading, thanks to the co-operation of many amateurs, considerable evidence showed that the half-wave aerial held the superiority, not only with respect to increasing the ground wave, but also in diminishing that type of fading due to interference between the sky wave and the ground wave.

The authors of the present paper plot Figs. 1 and 2 in terms of $H = 0{\cdot}55\lambda$, whereas they prove later on that $0{\cdot}56\lambda$ is nearer the optimum. Would it not have been more dramatic to have plotted Fig. 2 in terms of $H = 0{\cdot}56\lambda$ so as to underline this interesting concept of an optimum?

Towards the end of Section 1 the authors state that "the ground-wave field strength can be measured quite accurately . . ." The authors might have qualified this statement in the sense that field strength varies vastly within a very small area. As an illustration of this I can categorically state that a friend of mine living at the top of a tall building gets the Third Programme clearly, while I, living in a semi-basement, am denied it. If anyone should point out the obvious solution of a roof aerial and a down lead I would cite other difficulties not of a technical nature.

Mr. J. K. S. Jowett: The authors include some theoretical and measured assessments of the effects of siting the mast on a conical hill and on a circular plateau. They do not treat the additional case when there is a line of hills or a bump in the contour at points fairly near to the mast itself. Perhaps this situation does not lend itself to ready analysis, but I should be interested to know whether the authors can give some rough quantitative idea of the likely effects.

I can well imagine that many hours have been spent in trying to relate the ground contours shown in Fig. 21 to the measured results round Daventry. No reference is made in the paper to the authors' conclusions about the way in which these particular ground contours may have influenced the results, but it would be interesting to know whether the authors reached any conclusions concerning this problem.

Mr. Millington suggests that the ionosphere is the villain of the piece, and I do not question that as a general statement. On the other hand, were it not for the roughness of the ionosphere, perhaps the results shown in Fig. 26 would have been even more variable than they are. If one studies this Figure, it appears that, on the whole, large changes in field strength do not generally occur for bearing changes of less than 5° or 10°. Would the authors attribute this to the effects caused by a roughly reflecting ionosphere in blurring over some of the field-strength variations caused by the irregular ground?

I should like to say something, too, about the possible application of the evidence given in the paper to the synchronized operation of medium-wave transmissions. Clearly in the case of high-power broadcast transmitters, where a single channel is shared only by stations normally operating at spacings of at least several hundred miles, the interfering radiation from one station to the other is that which is transmitted at a low angle to the ground, and it seems that none of the arrangements discussed in the paper can materially alter this interfering radiation without also altering the service range of the station itself. But there are many cases of synchronized operation where networks are working with a common programme at close spacings. In this case, the radiation that causes interference is that transmitted at a considerably higher angle to the ground —something between 30° and 60° to the vertical. It appears from some of the measurements given in Figs. 18 and 19 that one could materially reduce the interfering radiation if the site chosen resembled the conical hill discussed in the paper. Do the authors think that this is a point which can, or should, be taken into account in the siting of medium-wave stations which are working in a synchronized network?

Mr. W. R. Piggott: In this classical paper the authors succeed in computing the properties of medium-wave aerials to a surprisingly high accuracy. It is therefore worth considering whether the rather limited published data on the reflecting properties of the ionosphere for these frequencies at night cannot be extended and put into forms applicable to this problem.

Pulse measurements at vertical and oblique incidence indicate that the main complexities can be readily resolved. At vertical incidence the absorption variation with frequency is dominated by the deviative absorption band near the critical frequency of the E-layer, which moves through the m.f. band during the evening and night. As soon as the propagation becomes oblique this anomalous absorption is confined to the F-reflection and the absorption for the E-reflection is mainly non-deviative, falling reasonably smoothly as the frequency and time after sunset increase. Thus in the fading zone the F-reflection is more absorbed than the E-reflection, as well as having a longer path, and the anti-fading problem is determined by the characteristics of the E-reflections. While the authors have given an excellent average value for the reflection coefficient of an E-reflection at night, it is clear that it will be too high on the lower frequencies and too low on the higher, and should increase regularly with time until the E-layer becomes partially reflecting.

The economic value of an anti-fading aerial depends partly on the periods for which it will prevent serious fading, and this may be estimated from the incidence of strong E-layer reflections. Although further work is desirable, we can say with fair certainty that in England this incidence varies, on the average, simply and regularly with time after sunset and season, but is little affected by changes in solar cycle. Thus a comparatively limited series of observations should be adequate to enable estimates of the relative value of anti-fading systems at different frequencies to be made.

THE AUTHORS' REPLY TO THE ABOVE DISCUSSION

Messrs. H. Page and **G. D. Monteath** (*in reply*): Mr. Millington refers to the notable contribution made by Böhm in calculating the feed current. In this calculation the primary current distribution, which is sinusoidal, is initially supposed to be maintained by an infinite number of elementary series generators, which supply the radiated power; the aerial is then assumed to behave as a dissipationless transmission line, carrying power from the base to the elementary generators. Clearly, on this

assumption the power flowing at the base must exactly equal the radiated power. Böhm's treatment therefore obtains the power radiated by the primary-current distribution exactly, even though the calculation of the feed current is approximate.

It is true that the reflected-wave field strength can be calculated by using the Fresnel reflection coefficient. Nevertheless, at the short range used in our experiments with models, the surface wave contributed appreciably to the field at angles to the vertical greater than 70°. We took the surface wave into account only in order to normalize the measurements for comparison with the theoretical curves.

We agree that interference (in the sense used in the paper) and fading both spring from the same cause. We tried to use commonly accepted terms, and at the same time to distinguish between a form of interference which can be reduced by increasing the transmitter power and one which cannot be reduced by this means.

We agree with Mr. Millington's comment on Section 2.2.1.

The investigation described in the paper was undertaken because the performance of some of the B.B.C. anti-fading aerials was not so good as had been predicted; the diagnosis has therefore come too late to effect a cure in many cases. The performance of the mast radiator at Daventry is better than if this work had not been undertaken. Another result has been the conversion of the Brookmans Park mast radiator to loop feeding. The main result, however, has been of a negative character, in helping us to decide what not to do. We feel, for instance, that further refinements of aerial design with the object of improving the vertical radiation pattern may not be worth while, because of the limitation mentioned in the paper; we now have the information available for deciding each case on its merits. It has not been possible to assess the reduction in fading from listeners' letters, since at Daventry the introduction of the new aerial was accompanied by a considerable increase in power, while the service area of the Brookmans Park aerial is restricted to some extent by foreign interference.

We are glad to have this opportunity to pay tribute to the important contributions made by Mr. Eckersley and his colleagues to our knowledge of broadcasting aerials. The paper he mentions describes the first experimental confirmation of Ballantine's theoretical work. Mr. Eckersley's own strong support in the early days of broadcasting for high aerials as a means of increasing the service area has been fully justified. Our paper is directed towards filling in some of the details, and explaining the failure to realize some of the earlier expectations of improved performance. In this connection it may be of interest to mention that the lack of agreement between theoretical predictions and measurements was the subject of comment at the discussion on Mr. Eckersley's paper. We consider that the theoretical calculations shown in his Fig. 7 present too rosy a picture of the performance of a high aerial, even under idealized conditions.

It might have appeared more logical to have plotted Fig. 2 for $h = 0 \cdot 56\lambda$, the optimum height, but we were reluctant to suggest that the optimum is so clearly defined as to be known to within $0 \cdot 01\lambda$. It must be remembered that our definition of the optimum height is somewhat arbitrary.

We agree that the ground-wave field strength may vary considerably from point to point within a small area. We were stressing the fact that, whereas the ground wave at a given point could always be measured, for instance during the daytime, it is much more difficult to determine the field strength of the reflected wave, which is a very variable quantity.

In reply to Mr. Jowett, our method of analysis is readily applicable to simple irregularities, such as a long ridge or valley, in the vicinity of the mast. Space does not permit any outline of the results, which tends to confirm our tentative definition of a "flat" site, given in Section 6.3 of the paper. We have attempted, by ruthless approximation, to calculate a curve corresponding to Fig. 26 from the ground contours shown in Fig. 21. Results so far obtained agree with Fig. 26 in respect of the mean field and its range of variation, but the point-by-point correlation between theory and experiment is nil. Mr. Jowett is probably correct in supposing that the curve shown in Fig. 26 has been smoothed by diffuse reflection.

We agree that, using a single high mast, it is not possible to reduce interference with a common-channel station several hundred miles away without also altering the service range of the station. It would be possible to go some way towards achieving this aim, at least in a limited number of directions, by using a number of aerials spaced horizontally; but we feel that the improvement would be limited by the effect of ground irregularity and possibly by diffuse ionospheric reflection.

In this country the high-power synchronized stations and the regional stations share the same site, the regional transmission usually employing an anti-fading mast radiator. It would be difficult to take advantage of ground irregularity to improve the vertical radiation patterns of both aerials. For low-power synchronized stations the area within which these can be sited is usually severely limited, and it would rarely be possible to obtain a site on a hill of suitable shape and high ground conductivity.

We agree with Mr. Piggott that information on the reflection coefficient of the ionosphere, expressed in a convenient form, would be of great value. Nevertheless we feel that the unpredictable component of the variations is greater than he suggests, so that observations over a long period would be required to obtain the data. It is surprising that progress in this field lags so far behind that in tropospheric propagation, which began to receive attention much later.

DISCUSSION ON

"A SHORT MODERN REVIEW OF FUNDAMENTAL ELECTROMAGNETIC THEORY"*

NORTH-WESTERN CENTRE, AT MANCHESTER, 2ND MARCH, 1954

Mr. L. H. A. Carr: In Section 3 the author raises the question of a dimensional constant to represent the concept of shape. Although I know that this is causing some heart-burning in I.E.C. circles, I feel most strongly that neither its discussion nor its use falls within the bounds of science or natural philosophy—that portion of knowledge where experiment is the criterion and test of credibility—but that it is a matter of concern only to the purely theoretical philosopher.

To the experimental physicist a volume is still a volume, whether its unit be taken as a cube of unit side or a sphere of unit radius; and to suggest anything different is only to create a further and unnecessary difficulty for students to overcome. There is just as much justification for including a further "dimensional constant" to indicate whether the ultimate standard is kept at Paris or in London.

The paper includes an attempt to reinstate the unit magnetic pole as the foundation of magnetostatics, but the author has succeeded in making his argument plausible only by ignoring the inconvenient characteristics of this theory.

For any mental concept to be acceptable, it must follow the same mathematical laws as does the natural phenomenon it represents. The field concept is one of the most useful auxiliary concepts, and I cannot understand the author's strictures about it, unless he considers that it includes some physical aspect that "pushes things around." As I, and many others, use it, it is simply a mathematical tool, enabling one to determine more readily, and evaluate effects due to, "the position and movement of charge," while it is entirely independent of whether those effects are due to "action at a distance" or not. The field concept thus only associates a mathematical equation with any particular point in space, without the postulation of any particular physical background.

By the application of field mathematics to the space surrounding first an isolated electric charge and secondly an isolated magnetic pole, the configurations are seen to be entirely dissimilar; in the former case there is spherical symmetry; in the latter the only symmetry is axial. Consequently eqns. (1) and (2) are not comparable, and the whole structure of a magnetostatics based on unit magnetic pole, in the same way that electrostatics can be based on a unit charge, falls to the ground.

In Section 7 the author writes: "By the choice of mass, length and time as basic dimensional quantities, F has been given the dimensions ML/T^2." But following his own line of argument, with which I entirely agree, there is no *a priori* reason why the two sides of the equation should be of the same dimensions, since the experimental data on which it is based refer only to the arithmetical figures. If the equation is to be used dimensionally, a dimensional constant must be included, although I agree that it is usual to suppress this and it is frequently very convenient to do so. It must, however, be realized that such a suppression is a human action, and any difficulties that arise as a consequence are not inherent in nature, but are of our own making.

Similarly, if, following the author, Q is adopted as a fourth basic dimension, the dimensional constant in the equation for force between charges can be suppressed or not, as we choose, the constant κ_0 being a pure numeric in the former case. In this case, however, the dimensional constant in the corresponding equation for force in magnetostatics (however expressed) must be retained, as is well known.

This reference to force leads to the question: What do we mean by force and how do we define it? It seems necessary to utilize this same equation for the purpose, and define force as something that has the power of bestowing acceleration on a mass. This, however, brings the concept of motion into the definition of a force, as does the alternative definition of the space rate of doing work. It is therefore very doubtful whether force can be considered without relation to motion or change of motion. In this case the simple formula for the force on a conductor carrying current in a stationary magnetic field cannot justifiably be applied to a conductor rigidly held in an armature slot, the magnetic field surrounding which undergoes change while the conductor is in motion.

I have never seen any satisfactory basis put forward for the hypothesis repeated by the author in Section 14 that the mechanical forces in a machine with slotted armature "act largely on the iron teeth and not on the conductors in the slot," and I suggest that the author's theories, where he states that "the balance [of energy] can be achieved only by a transformer effect which results in an induced e.m.f. in the coil," if worked out in full detail, would show on the basis of his method of calculation that there was a mechanical force between the conductor and the teeth, resulting in the whole of the force finally coming on the conductor, albeit transmitted to the field-magnet poles through the medium of the armature teeth.

Professor E. Bradshaw: "Preoccupation with the doctrine of flux" may by some be thought to be undesirable, but such an attitude does not necessarily follow from the adoption of the M.K.S. rationalized system of units. The author admits that, in support of the concept of the unit magnetic pole, . . . "we impose symmetry"; the symmetries invoked by those who emphasize the field concepts are surely no more artificial. Apart from the welcome, on educational grounds, given to the M.K.S. system by the author, it should be stressed that any teaching sequence can be adopted using this or any other consistent unit system.

In Section 8, in connection with rationalization, the author says that the student "should not be allowed to make his choice until . . ." This is surely a counsel of perfection and assumes a remarkable type of student. Would it not be more realistic to suggest that the student, having first acquired a satisfactory understanding of the basic relations of electrical science via whatever path the teacher deems to be most direct and consistent, may then be exposed to other systems of units and relations?

Mr. E. Wild: The fundamental formula assumed in Part 2 for the induced e.m.f. is eqn. (14), e.m.f. = − rate of change of flux. The separation of the rate of change of flux into a transformer part and a motional part [eqn. (22)] is obtained from this by a mathematical transformation. If the two formulae gave different results in any case, there would be a self-contradiction in the theory which would not be removed by the author's

* HAMMOND, P.: Paper No. 1595, December, 1953 (see 101, Part I, p. 147).

device of choosing one formula or the other as correct. It is therefore necessary to reconcile the apparent contradictions.

The fundamental formula applies to continuous motions of linear circuits, i.e. circuits composed of wires of negligible cross-sectional dimensions. If the motion is discontinuous, or if the circuit contains elements of large cross-sectional dimensions, such as the conducting strip in Fig. 8 or the magnet, M, in Fig. 5, through which an infinite number of conducting paths can be drawn, the formula must be supplemented by further conditions, namely

If the circuit is changed discontinuously, e.g. by switching, the resulting discontinuous change of flux produces no e.m.f.

The law of induction is to be applied only to circuits every part of which moves with the material in which it lies, sliding contacts being permitted.

With these conditions it can be shown that, in all the cases described in Section 15 in which the fields are constant, the flux-cutting and flux-threading rules are equivalent.

The effect of finite cross-sectional dimensions alone is seen in the experiment depicted in Fig. 8. If the circuit through the galvanometer is completed by any line crossing the strip, and moving with the strip, together with the edges of the strip, it can be seen that the flux-threading rule applied to this circuit gives the same results as the flux-cutting rule.

The effect of switching alone is seen in the experiment of winding a coil on to a ring magnet with negligible leakage field by means of a circular slip-ring surrounding the magnet; one end of the coil is attached to the slip-ring at a point A, and the end of the unwound wire has sliding contact with the slip-ring at point C. As the ring rotates, turns are wound on the coil but no e.m.f. is generated. Here there are two circuits to be considered: those that complete the circuit of the coil through the slip-ring by going from C to A in the sense of rotation and in the opposite sense. The flux through either circuit remains constant as the slip-ring rotates, except when A passes C. Consider the circuit completed by the arc of the slip-ring which goes from C to A in the sense of rotation. The geometrical configuration of the circuit specified in this way changes discontinuously as A passes C, the length of the arc of the slip-ring involved changing from the full circumference to zero. The change cuts out a turn (consisting of the slip-ring itself) wound in the reverse direction, so that the flux increases discontinuously by the amount corresponding to one turn of the coil. Thus the flux through the coil increases discontinuously once per revolution, but the rate of change of flux is always zero, and so is the e.m.f.

The other examples can be explained by a combination of these two methods.

The new derivation of the electromagnetic equations proposed in Part 3 contains a fundamental flaw. Eqn. (41) and eqn. (43), which is derived from it, are self-consistent only if the vector field $\boldsymbol{D}$ satisfies certain conditions, the most general form of which implies that the electric charge has everywhere a definite velocity, the velocity $\boldsymbol{u}$ which occurs in the equations. The equations are thus possible ones for the hydrodynamics of charged fluids in appropriate conditions, but are not adequate for general electromagnetic theory.

The equations, in fact, do not agree with the generally accepted ones, since the current density is not in general, $\boldsymbol{u}$ div $\boldsymbol{D}$. In most engineering applications current flows in electrically neutral media where div $\boldsymbol{D}$ is zero. Eqn. (45) would imply that an uncharged conducting wire carrying a current generates no magnetic field.

Mr. H. B. Daniels: The terms "cutting" and "threading" are both used in considering electromagnetism, but I find the idea of "linkage" (of electric and magnetic circuits) very valuable in covering the needs of conduction and radiation of electrical energy. The simplest case to visualize is a circular copper conductor linked with an iron ring, a current in the conductor producing magnetism in the ring, and change of magnetic flux, with which the conductor is linked, producing an e.m.f. in the conductor. This is useful for introductory ideas for material circuits (conduction) and for space (electromagnetic radiation). In the latter case we can imagine the materials removed, and there are still electromagnetic effects in the field of radiation; conduction may be considered as a special case of electromagnetic radiation. Moreover, the wavelength of the electromagnetic propagation is dependent on the frequency of the original electrical oscillation (e.g. in an aerial). The current circuit is based on the physical reality of a quantity of electricity in motion, and thus, with the fourth fundamental quantity accepted as quantity of electricity, we have an M.K.S. system of units unifying conduction and radiation concepts.

Mr. W. E. Burnand (*communicated*): I question whether linkage, cutting or detached magnetic poles form the complete story of electromagnetic induction; I think that there is also another component.

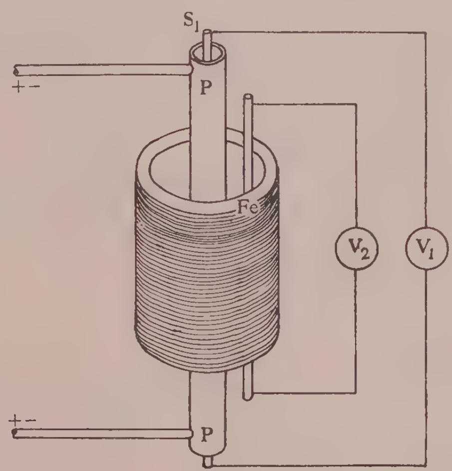


Fig. D

Consider Fig. D. P is a copper tube and S_1 is a conductor inside P. S_2 is a conductor outside P but inside a substantial laminated iron cylinder Fe. It is well known that no magnetic field is generated inside a tube carrying a uniformly distributed current longitudinally.

If now a current is passed along P, a magnetic field is generated round it (but not inside), proportional to the current in the tube and the magnetic permeability of its surroundings, and inversely proportional to the radial distance. With an alternating current along P of such a value that Fe is not saturated, the field in the air spaces is small and may be neglected for the purpose of the present discussion, the significant field being that in the laminated iron cylinder. The voltage induced in P, S_1 and S_2 by the varying magnetic field of Fe is the same. In the space between P and Fe there is the m.m.f. spreading out from P and the small magnetic flux, but inside P there is neither m.m.f. nor magnetic flux; but there is still the same voltage induced in S_1. As the radiation from P is all outwards, it follows that the induced voltages are due, not to this outward radiation, but to an inward radiation from the varying magnetic field, which, moreover, is of a different character to the outward radiation, since it has neither m.m.f. nor magnetic flux.

Normally the whole of the induced voltage can be considered as confined to the space surrounded by Fe, and in fact is so, as evidenced by the performance of innumerable current transformers in daily use. But the possibility arises that there might be some way in which this non-magnetic voltage-inducing component could be projected beyond the plane of the generating field, analogous to the way in which the electromagnetic wave is projected in the usual radio transmission.

If the magnetic field round a series of copper conductors is explored as these are progressively added till they form an elongated cage, it will be found that as these reach a circle they form the equivalent of a tube, in that the magnetic field goes completely outside the circle leaving the inside clear of magnetic flux; but the full inductive effect remains.

A large circle of vertical conductors, fairly close together but of the dimensions of a Druid circle, caused to oscillate in unison under crystal control, looks an attractive experiment, since this would radiate the usual electromagnetic wave outwards from the circle and the different, voltage-inducing non-magnetic component from the inside. Since this latter component is created in any case by a more or less circular varying magnetic field, it seemed worth trying a fairly large toroidal coil, in spite of the fact that the non-radiating property of the toroid is so universally recognized and made use of in radio apparatus.

A toroid of about 30in diameter was therefore constructed, using a piece of rubber hose bent into a circle and wound uniformly with about 2 000 turns of 22-gauge copper wire. When a small current (about 0·4amp) was passed at about 2kc/s by means of a valve of unknown characteristics, the characteristic squeal was readily picked up at a distance of 7ft by means of a search coil and amplifier, the coil being about 7in in diameter and having 64 turns, the assembly being as indicated in Fig. E.

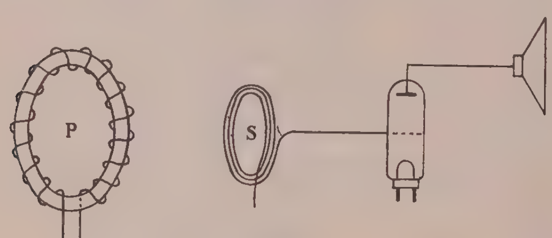


Fig. E

Another search coil of nearly the same diameter as the toroid gave similar results. The loudest signal was received with the toroid transmitter and straight-coil receiver squarely facing. With the toroid turned through 90° so as to face the receiving coil edgewise, or the search coil turned 90° facing the transmitter edgewise, the signals ceased.

Thus we have transmission through space with no magnetic interlinkage, no magnetic flux or m.m.f. projected from the toroid, no "cutting" or varying "linkage" of magnetic flux, and the conductors in the transmitting toroid at right angles to those in the receiver or secondary circuit.

Static can be ruled out, as signals ceased when the toroid winding was opened at its mid-point. Transmission was effective throughout the audible range of frequency and also when the toroid was fed from a 500c/s alternator about 40ft away as a check on whether signals were picked up from the valve oscillator. There is still left the third component or element as the transmitting agent.

Dr. R. Feinberg also contributed to the discussion at Manchester.

Mr. P. Hammond (*in reply*): Mr. Carr objects to the use of the concept of pole strength, because the field around an electric charge exhibits spherical symmetry, but around an isolated pole there is axial symmetry. If this were indeed so, I should be the first to abandon the concept of pole strength. But Mr. Carr's statement is equivalent to saying that the law of force between poles is not that of the inverse square. Because the laws *are* identical for electric charges and magnetic poles, their field patterns are identical. In fact, the field around an isolated charge and the field around an isolated pole both exhibit spherical symmetry. This is the justification for the invention of the twin concepts of a point charge and a point pole. It would hardly be an exaggeration to say that this approach halves the mental effort required from the student. I agree with Professor Bradshaw that the M.K.S. system need not be tied to a particular sequence of instruction. It is all the more regrettable that the over-enthusiastic supporters of the system wish to give it a philosophical content that it does not possess by tying it to the Maxwellian aether theories. The subsidiary conditions that Mr. Wild attaches to Faraday's law will undoubtedly give the correct answer, but to some students the conditions may appear unconvincing. For instance, how can the rate of change of flux be always zero and yet the flux be increasing? If the flux increases discontinuously, will not its rate of change be infinite? Surely all these conditions are unnecessary, if it is realized that eqn. (18) is identical with eqn. (22). I would urge Mr. Wild to re-read Section 16 of the paper. With regard to eqns. (41) and (43), div $\boldsymbol{D}$ is not zero in a conductor. Every electron implies that there is a divergence of $\boldsymbol{D}$.

I find Mr. Burnand's contribution very difficult to understand. The success of Maxwell's theory has been such that it seems unlikely that a completely new type of radiation would have to be postulated to account for Mr. Burnand's experimental evidence. The trouble experienced by Mr. Burnand seems to arise from a too ready acceptance of such statements as that contained in his second paragraph. There is, in fact, always a magnetic field inside a tubular conductor carrying alternating current, in spite of the widespread belief to the contrary. Similarly, there is a magnetic field outside a solenoid. Lack of space prevents a detailed analysis here, and I would refer Mr. Burnand to Professor E. B. Moullin's book "Radio Aerials," where these problems and many similar ones are treated in detail.

621.317.725.082.72

Paper No. 1675 M
July 1954

AN ATTRACTED-DISC ABSOLUTE VOLTMETER

By G. W. BOWDLER, M.Sc.

(*The paper was first received* 16*th January, and in revised form* 25*th March*, 1954. *It was published in July*, 1954, *and was read before the* MEASUREMENTS SECTION 14*th December*, 1954, *and the* NORTH-WESTERN CENTRE 3*rd May*, 1955.)

SUMMARY

Details are given of the design and performance of an attracted-disc voltmeter, operating in a medium of compressed gas, which is capable of measuring the effective value of a wide range of voltages. The accuracy with which measurements can be made increases with the magnitude of the voltage and is about 1% at 1 kV and 0·1% at 30 kV and above. The clearances between the electrodes have been designed to withstand voltages up to 700 kV (peak), but measurements have been limited to 350 kV (peak) by the insulator used to support the high-voltage electrode.

LIST OF PRINCIPAL SYMBOLS

a = Width of annular gap between disc and guard plate, m.
A = Effective area of attracted disc, m^2.
c = Distance between guard plate and h.v. electrode, m.
E = Mean electric field between the electrodes ($= V/c$), volts/m.
E_0 = Electric field at centre of disc, volts/m.
E_s = Electric field at distance s from centre of disc, volts/m.
F = Force acting on the disc when it is displaced by distance h, newtons.
F_0 = Force acting on the disc when it is coplanar with the guard plate, newtons.
F_E = Elastic force tending to make the disc coplanar with the guard plate when it is displaced by distance h, newtons.
g = Acceleration due to gravity, m/s^2.
h = Height of disc above plane of guard plate, m.
m = Mass of brass weight, kg.
r = Effective radius of attracted disc, m.
V = Effective value of applied voltage, volts.
ϵ_0 = Permittivity of free space ($= 8{\cdot}854 \times 10^{-12}$ farads/m).
ϵ_r = Permittivity of medium between electrodes relative to that of free space.
ρ_1 = Density of brass weight, g/cm^3.
ρ_2 = Density of gas medium, g/cm^3.

(1) INTRODUCTION

The force of attraction between two electrified bodies has long been used as the basis of voltage measurement. Lord Kelvin[1] was the first to design an instrument in which the relation between the force and the magnitude of the voltage in absolute electrostatic units could readily be calculated; he used parallel plane electrodes, and in order to avoid complications due to stress concentration at the edges, he measured the attractive force on a disc forming the central portion of one of the electrodes and separated from the remainder by a narrow annular gap. He showed that the effective area of the disc was very closely equal to the mean between its actual area and the area of the hole in the surrounding guard electrode.

During the last few decades, with the introduction of ever-increasing transmission voltages, several instruments based on Lord Kelvin's design have been constructed for the measurement of high voltages. Of these, the most notable was that described by Brooks, Defandorf and Silsbee of the National Bureau of Standards[2] which was capable of measuring, to an accuracy of a few parts in 10^4, voltages of sine waveform up to 275 kV (r.m.s.). The electrodes of this instrument were not enclosed, and special precautions were necessary to avoid disturbances due to draughts. Several German workers,[3,4] on the other hand, have designed instruments enclosed in a chamber filled with compressed gas, thereby overcoming difficulties due to draughts and deposition of dust on the electrodes and allowing the voltage gradient, and hence the attractive force between the electrodes, to be considerably increased; in this way an instrument operating up to 400 kV (r.m.s.) has been constructed.[4] The instrument about to be described is a Kelvin-type absolute voltmeter, using a medium of high electric strength, for the measurement of a wide range of voltages.

(2) GENERAL DESIGN CONSIDERATIONS

As a basis for the design, it was assumed that an instrument capable of measuring, to an accuracy of 0·1%, voltages up to 500 kV (r.m.s.) was required.

Three suitable media having a high electric strength are transformer oil, high vacuum and compressed gas. Oil was rejected on account of the influence of contaminating particles on the electric strength and the fact that the two alternative media are able to sustain electric stresses as great as those of well purified oil. Of the other alternatives, a medium of compressed gas was preferred to one of high vacuum because of the ease of dealing with gas leaks in the system.

The instrument thus consists essentially of a pair of parallel-plate electrodes in a pressure vessel, and facilities are provided for measuring the attractive force acting on a disc forming the central portion of one electrode, for detecting displacement of the disc from the plane of the surrounding guard plate and for varying and measuring the distance between the electrodes. It is convenient for the disc, guard plate and pressure vessel to be earthed.

Using the rationalized M.K.S. system of units and assuming a uniform field between the electrodes, the attractive force F_0 on the disc when it is coplanar with the guard plate is given by the equation

$$F_0 = \frac{\epsilon_0\epsilon_r AE^2}{2} \quad \text{or} \quad \frac{\epsilon_0\epsilon_r\pi r^2V^2}{2c^2} \quad . \quad . \quad . \quad . \quad (1)$$

(2.1) Size of Electrodes

Published data[5] on the electric strength of compressed gases indicate that stresses up to 100 kV (r.m.s.)/cm can be sustained by air, nitrogen or carbon dioxide at 15 atmospheres pressure and by dichlor-difluor-methane (Arcton 6) or sulphur hexafluoride at 3 atmospheres pressure. For these media a maximum electrode separation of about 5 cm is therefore necessary, and the high-voltage and guard electrodes should be of sufficient diameter to ensure that the electric stress over the surface of the disc is uniform to a high degree at this separation. The h.v. electrode should also be well rounded at the edge so that at no point on its surface does the stress greatly exceed that existing

The paper is an official communication from the National Physical Laboratory.

between the electrodes. With an assumed disc diameter of 10 cm, rough calculations of electric stress derived from formulae relating to parallel-plate and concentric-cylinder electrodes gave the following dimensions for the electrodes and the containing vessel:

Overall diameter of h.v. electrode	0·4 m
Radius of curvature of edge of h.v. electrode . .	0·04 m
Diameter of guard plate and of interior of tank	0·56 m

The maximum value of the attractive force, calculated by putting $r = 0{\cdot}05$ m, $E = 10^7$ V/m, and $\epsilon_r = 1$ in eqn. (1), is

$$F_0 = 3{\cdot}5 \text{ newtons}$$
$$\simeq 0{\cdot}35 \text{ kg or } 0{\cdot}77 \text{ lb}$$

(2.2) Stability of Attracted Disc

The disc should be mounted on an elastic support so that it is normally coplanar with the guard plate and separated from it by a narrow annular gap and is free to move in a direction perpendicular to its plane. On the application of a voltage V to the electrodes, a force F will arise attracting the disc towards the h.v. electrode, and this must be balanced by a measurable force acting in the opposite direction, balance being determined by the restoration of the disc to its normal position.

The stiffness of the elastic support and the sensitivity of the detector of displacement of the disc should be such that balance of the opposing forces can be carried out to the required degree of precision and such that there is no instability of the disc. Brooks, Defandorf and Silsbee[2] have investigated the variation of the attractive force F on the disc with its displacement h from the plane of the guard plate. When h is small compared with the width a of the gap between the disc and guard plate,

$$F = F_0(1 + fh) \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where $$f = \frac{2}{c} + \frac{2}{\pi r}\left\{2 \log_\varepsilon\left[\frac{2c(1-\varepsilon^{-\pi r/c})}{a}\right] - 1\right\}$$

Hence, when $fh \ll 1$, $$\frac{1}{F}\frac{dF}{dh} \simeq f \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

The incremental force dF acts so as to make the disc unstable, and unless the elastic restoring force dF_E, called into play by the displacement dh, exceeds dF the system will actually become unstable.

If $a = 0{\cdot}0003$ m and $r = 0{\cdot}05$ m,

$$f = 310 \text{ m}^{-1} \text{ when } c = 0{\cdot}01 \text{ m}$$

and $$f = 190 \text{ m}^{-1} \text{ when } c = 0{\cdot}05 \text{ m}$$

and with a value of $F_0 = 0{\cdot}35$ kg (Section 2.1), the maximum value of dF/dh will be $0{\cdot}35 \times 310 = 110$ kg/m. The elastic constant dF_E/dh of the disc support must therefore not be less than 110 kg/m.

(2.3) Sensitivity of Detector of Displacement of Disc

If, because of insufficient sensitivity of the detector of displacement during the balancing operation, the disc is not adjusted exactly to its normal coplanar position, two sources of error are introduced. The first of these is a change dF in the attractive force from its value F when the disc is coplanar with the guard plate, resulting in a fractional error $dF/F = fdh$ [eqn. (3)]. For a given displacement dh the fractional error therefore depends on the value of f, which increases as the electrode separation decreases but, for the electrode dimensions quoted in Section 2.2, only changes from 190 to 310 m^{-1} as the electrode separation changes from 5 cm to 1 cm. Thus, if $dh = 1$ micron (10^{-6} m) the fractional error dF/F varies over the range 2–3 $\times 10^{-4}$.

The second source of error is the elastic restoring force dF_E brought into play by the displacement dh. If again $dh = 1$ micron and the stiffness of the disc support is no greater than the value 110 kg/m, required to ensure stability (Section 2.2), dF_E will be 0·11 g. Thus, with a displacement dh of 1 micron and a total force F of 0·35 kg (corresponding to a voltage gradient of 100 kV/cm), the fractional error dF_E/F will be 3×10^{-4}. This error will vary inversely as the magnitude of the force F.

The two sources of error due to maladjustment of the disc

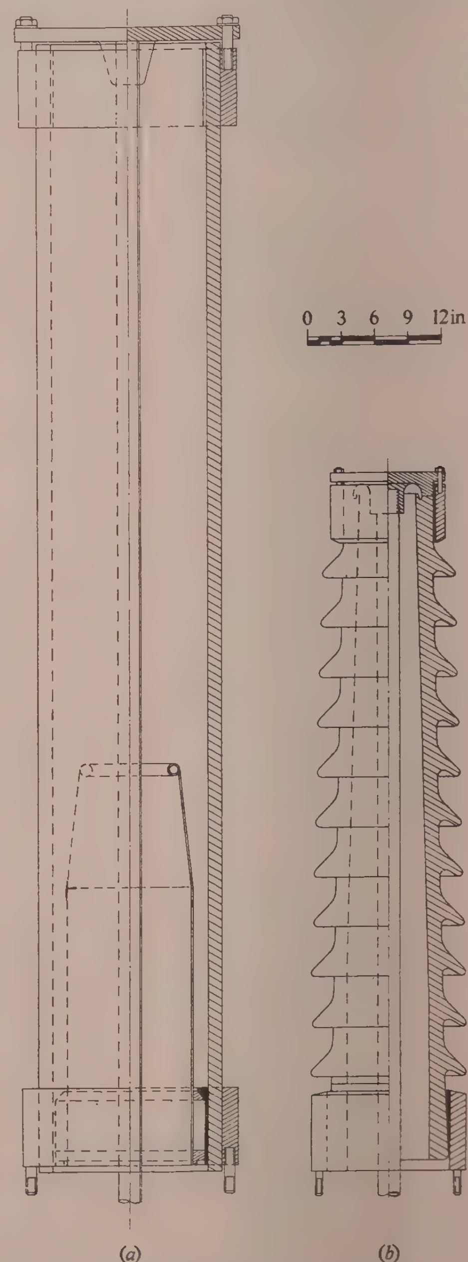


Fig. 1.—Insulators for h.v. electrode.
(a) Proposed design for 500 kV (r.m.s.).
(b) Porcelain cable sealing end.

are of opposite sign and will, in fact, just cancel each other when the instrument is used on the threshold of instability. With electrode spacings of 1–5 cm and voltage gradients of approximately 100 kV/cm between the electrodes, the above considerations show that a displacement detector sensitive to 1 micron will ensure that the attractive force F can be balanced to within about 1 part in 10^4; the uncertainty in balance will increase as the voltage gradient decreases and amount to about 30 parts in 10^4 at a gradient of 30 kV/cm.

(2.4) Insulator for High-Voltage Electrode

Experience has shown that a resin-bonded paper tube with metal end fittings and a tapered internal electrode, a general drawing of which is shown in Fig. 1(*a*), forms, when filled with compressed gas, a satisfactory bushing insulator for 500 kV (r.m.s.). Such an insulator has not yet been obtained; as a substitute a porcelain cable-sealing-end designed for a working voltage of $132/\sqrt{3}$ kV, a drawing of which is shown in Fig. 1(*b*), has been used.

(3) DESCRIPTION OF INSTRUMENT

An elevation of the essential parts of the instrument is shown in Fig. 2. An angle-iron framework carried on large castors supports a square steel-plate $1\frac{1}{2}$ in thick which forms the lid of a pressure tank having an internal diameter of 24 in and a depth of 26 in. A circular flanged turret in the middle of this plate

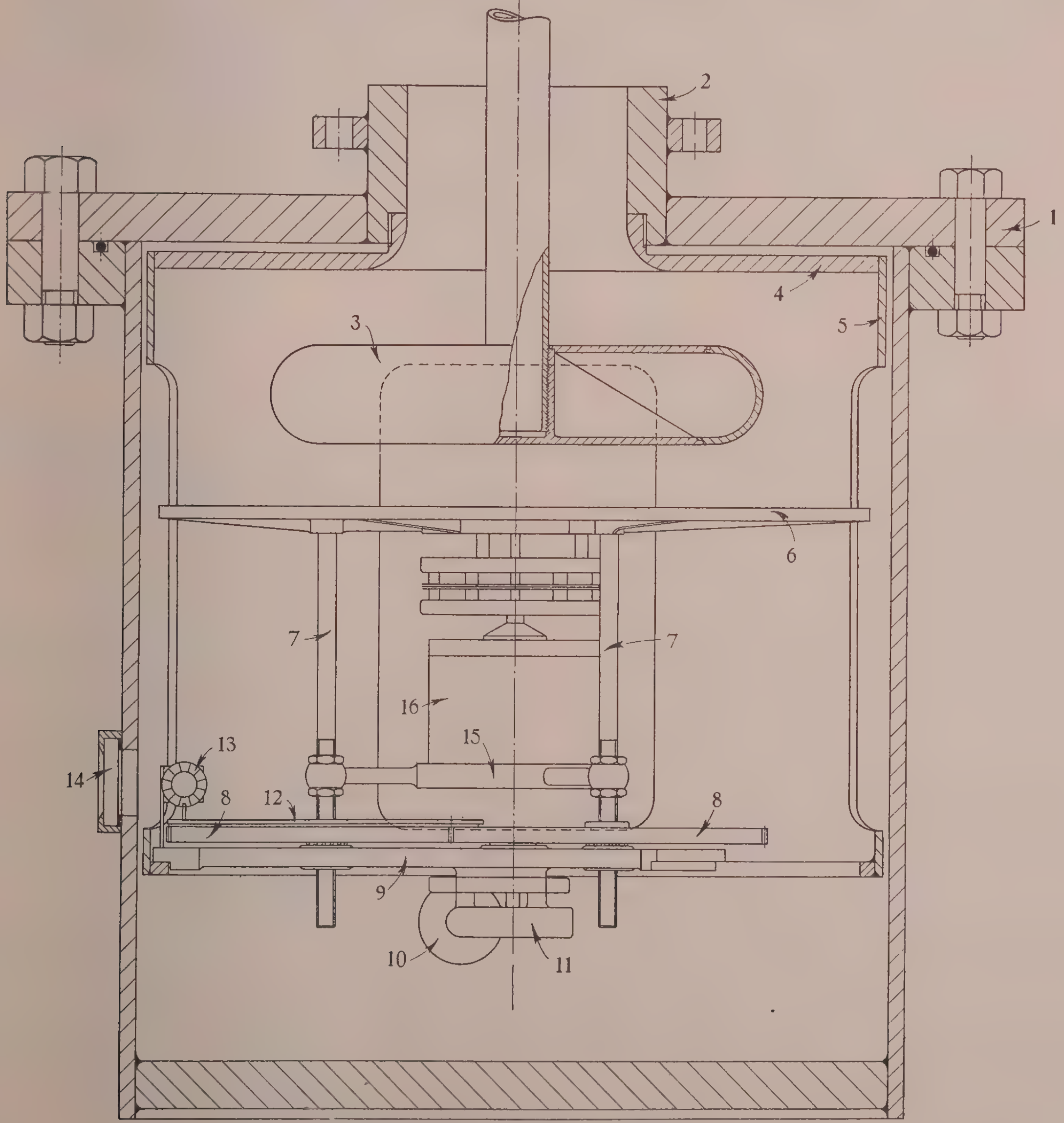


Fig. 2.—General elevation of instrument.

1. Square top plate.
2. Turret.
3. H.V. electrode.
4. Steel disc screwed to bottom of turret.
5. Thin-walled cylinder.
6. Guard electrode.
7. Screwed pillars supporting guard electrode.
8. Gear wheels.
9. 3-armed base casting.
10. Motor.
11. Worm gear box.
12. Micrometer scale.
13. Cyclometer counter.
14. Window.
15. Magnet support.
16. Magnet.

forms a support for the h.v. insulator, from the upper end of which the h.v. electrode is suspended by means of a steel tube of 2 in outside diameter, screwed at each end with a 1¼ in B.S.P. thread.

The h.v. electrode consists of a hollow bronze casting 16 in in diameter and 3 in thick machined all over with the edges rounded to a radius of curvature of 1·5 in. To obtain the requisite flatness, the under surface of the electrode is provided with internal ribs radiating from the central screwed boss by which the electrode is attached to its stem. After machining, the surfaces of the electrode and its stem were chromium plated and polished and the under surface was lapped. The weights of the electrode and stem are 23 kg and 17 kg respectively.

To avoid variation of electrode separation due to flexure of the tank lid when internal pressure is applied, the l.v. electrode and all its associated measuring equipment are supported from the underside of the turret via a steel disc, ½ in thick and 22½ in in diameter, screwed to the throat of the turret and a thin-walled cylinder reaching almost to the bottom of the tank. Four large rectangle apertures cut in the walls of this cylinder give access to the mechanism within. The junction of the surfaces of the steel disc and the bore of the turret is rounded to avoid a high concentration of stress in that region, and the surfaces themselves are nickel plated and worked to a smooth finish.

The guard electrode is a bronze casting, ribbed on the underside to give it rigidity, resting on three steel pillars. These pillars have accurately-formed screw-threads of 1 mm pitch cut on their lower end on which nuts turned by large-diameter (200-tooth) gear wheels are mounted. The lower ends of the pillars pass through holes in a 3-armed bronze base casting; the extremities of these arms rest on a ½ in square-section ring welded to the bottom edge of the thin-walled cylinder. Thrust ball-races are interposed between the nuts and the arms of the base casting against which they bear. By individual adjustment of the three gear wheels, the guard electrode can be tilted until it is accurately parallel with the h.v. electrode.* A common centre pinion, driven through a worm reduction gear by a small electric motor mounted on the base casting, can then be meshed to all three gear wheels and thus enable the electrode separation to be varied. A circular scale, mounted on one of the gear wheels, in conjunction with a cyclometer counter, enables changes in the electrode separation to be read through a small window in the wall of the pressure tank to an accuracy of better than 0·001 cm. The scale and cyclometer dial can be illuminated by a small internal lamp.

The attracted disc and its associated mechanism for measuring the force of attraction, which forms a separate unit mounted on the underside of the guard plate, is illustrated by the drawing shown in Fig. 3; its framework consists of two pairs of bronze rings AA and BB, and a smaller ring C, connected together by brass screws and bronze brackets and distance pieces. A bracket fixed to the ring B carries a lever pivoted at D and loaded at one end by a circular-shaped 100 g brass weight E which is rather more than counterbalanced by a brass block F on the other end. By energizing the solenoid G the weight F may be slightly lifted and E correspondingly lowered.

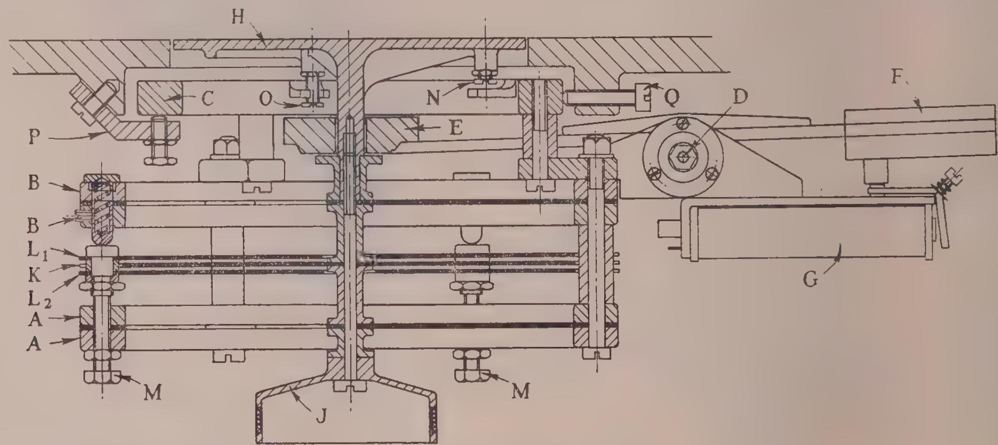


Fig. 3.—Details of attracted disc and its associated mechanism.

The light-alloy attracted disc H is connected by a long clamping screw, on which are mounted several flanged distance pieces, to a circular coil-former J, also of light alloy, wound with 430 turns of No. 40 S.W.G. enamelled copper wire; between one pair of distance pieces a circular duralumin plate K is clamped. This system is mounted on the supporting framework by two pairs of phosphor-bronze strips, which are clamped at one end between rings AA or BB and at the other end between distance pieces on the disc-system assembly screw. To obtain precision in the positioning of the disc, the components of each pair of phosphor-bronze strips are arranged with their axes at right-angles; the ends which are clamped together by the comparatively slender screw connecting the coil former to the disc are also reduced in thickness so that they exert no great bending moment on this screw. The four cantilevers supporting the disc system are thus approximately freely loaded at their ends. The disc system is about 190 g in weight and was initially designed to be mounted on two thin elastic diaphragms clamped between rings AA and BB, but these and also a pair of phosphor-bronze strips, spanning the rings and clamped by them at each end, were found to be much less satisfactory forms of support than those already described and finally adopted.

The circular plate K carried by the disc is situated between similar plates L_1 and L_2 mounted on the screws M, which are adjusted so that when the disc system is in equilibrium the three plates are parallel and the gaps (nominally 1 mm) between the middle and each of the outer ones are equal; this is achieved by applying an audio-frequency voltage between the middle and

* Owing to its weight and the length of its supporting stem the h.v. electrode is not very rigidly supported, so that the parallelism of the electrodes is dependent to some extent on the level of the instrument. With electrodes adjusted parallel when the instrument was level, it was found that a tilt through a small angle θ caused the electrodes to deviate from parallelism by an angle of approximately 0·015θ.

outer plates and then adjusting the screws M until an equal-ratio Schering bridge incorporating the capacitances between electrode K and each of the electrodes L_1 and L_2 is just balanced. The flange of one of the distance pieces on the disc assembly screw serves as a platform on which the weight E can be lowered. The disc and the plate K which moves with it are insulated from each other and from the frame; the plates L_1 and L_2 are also insulated from each other and from the frame. Screws N, N on the underside of the disc are adjusted to engage against stops projecting on the inside of the ring C when the disc is moved about 0·1 mm downwards from its normal position of rest; two other screws O, O at opposite ends of a diameter at right-angles to that through N N similarly limit the upward movement of the disc from its equilibrium position.

When all the necessary adjustments have been made on it, the disc unit is mounted on the underside of the guard plate by means of the ring C which rests on three equispaced brackets P provided with levelling screws. These screws are adjusted until the disc, when tested with a knife-type straight-edge, is coplanar with the guard plate; ring C is then clamped by three equispaced radial screws Q, which are adjusted so that the disc is concentric with the hole in the guard plate which surrounds it. The balance of the equal-ratio Schering bridge is again checked so that in all further use of the instrument it can be assumed that when the bridge is balanced the surfaces of the disc and guard plate are coplanar. Finally, a permanent magnet mounted on a 3-armed bronze support, which, in turn, is mounted on the three pillars supporting the guard plate, is adjusted vertically and laterally until the coil is situated centrally in the annular gap between the poles. The tank is then raised by means of jack screws operating in three of the bolt holes in its flange and bolted to the lid with a rubber O-ring interposed to make a gas-tight seal.

The instrument is connected, via a multicore cable and a multi-way plug and socket connector in the wall of the tank, to a control box which is fed from a 50-c/s supply and provides an audio-frequency source and visual detector for the Schering bridge, and d.c. supplies for the motor, solenoid, lamp and coil. A circuit diagram of the control box is reproduced in Fig. 4.

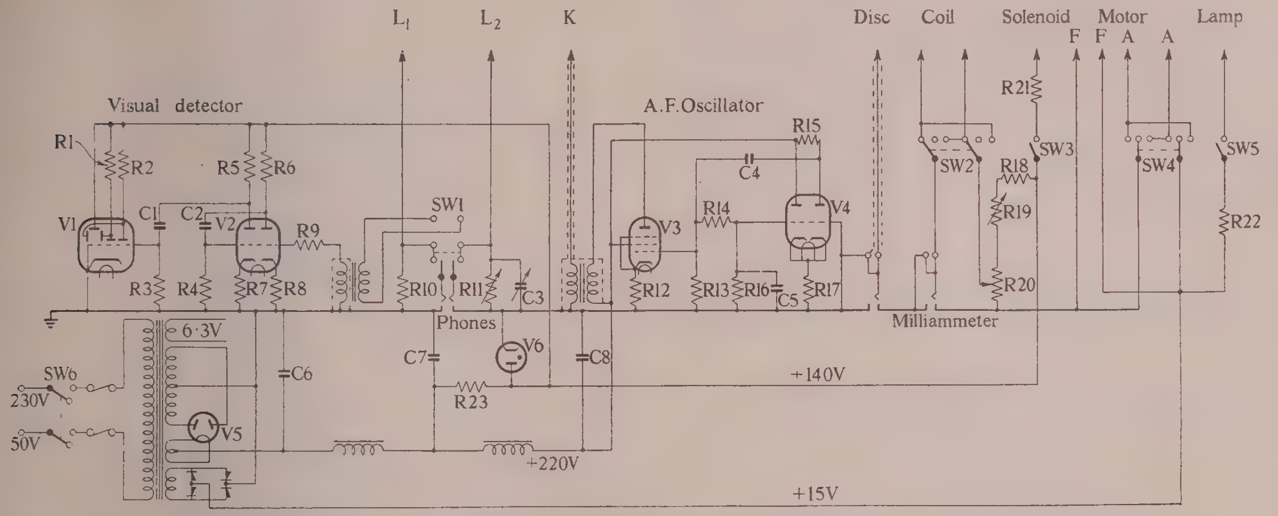


Fig. 4.—Circuit diagram of voltmeter control box.

V1	Electron-beam indicator, type EM34.	SW3	Single-pole solenoid switch.
V2	Double triode, type 12 AX7.	SW4	2-pole 3-way motor switch.
V3	Output pentode, type 7 D10.	SW5	Single-pole lamp switch.
V4	Double triode, type 12 AU7.	SW6	2-pole 2-way main switch.
V5	Full-wave rectifier, type 5Z4G.		
V6	Voltage stabilizer, German type.		
SW1	2-pole 2-way bridge detector switch.		
SW2	2-pole 3-way coil current switch.		

R1, R2, R3, R4	1 MΩ, 0·5 W	R11	Wire, variable in steps of 1 Ω from 1 950 to 2 050 Ω
R5, R6	500 kΩ, 0·5 W	R12	100 Ω, 0·5 W
R7, R8	2 kΩ, 0·5 W	R13	300 kΩ, 0·5 W
R9	5 kΩ, 0·1 W	R14	750 kΩ, 0·1 W
R10	2 kΩ, wire	R15	150 kΩ, 0·5 W
		R16	27 kΩ, 0·1 W
		R17	1 kΩ, 0·5 W
		R18	2 kΩ, 10 W, wire
		R19	10 kΩ, 50 W, wire

R20	1 kΩ, 2 W, wire
R21	3 kΩ, 5 W
R22	50 Ω, 5 W
R23	1·8 kΩ, 20 W, wire
C1, C2	0·001 μF
C3	200 μμF
C4	0·002 μF
C5	0·003 μF
C6, C7, C8	8 μF

(4) PROCEDURE IN USING THE INSTRUMENT

The complete assembly and adjustment of the instrument described in Section 3 is a somewhat lengthy process, but when it has once been done the instrument is very convenient to use and has proved to be trouble-free.

The reading of the micrometer scale corresponding to zero separation of the electrodes is first determined. This may be done either by raising the earthed-electrode system until contact between it and the h.v. electrode is just established and then noting the micrometer reading, or by measuring over a range of small values of electrode spacing the capacitance C between the disc and the h.v. electrode, the coil current being adjusted so that the disc is always in the plane of the guard plate. In the latter case the relation between $1/C$ and the micrometer reading is extrapolated to $1/C = 0$, the micrometer reading corresponding to this is the effective zero of the scale of electrode separation; it is found in general to differ from the contact zero by about 0·02 mm. Owing to differences between the coefficients of thermal expansion of the insulator and steel tube by which the h.v. electrode is supported, the effective and contact zeros of the instrument are dependent on the ambient temperature. For a particular set-up of the instrument the difference between the effective and contact zeros is constant, so that when this difference has been determined, the effective zero can readily be deduced from the contact zero.

After the effective zero has been obtained, the disc is earthed to the tank and the electrode spacing is increased to a value

appropriate to the voltage to be measured. The voltage is then applied, thus destroying the balance of the Schering bridge. Balance is restored either by lowering the brass weight of mass m on to the platform on the disc stem and then adjusting the electrode spacing, or by adjusting the current in the coil to a value I_1. In the first case the force on the disc is equal to $mg(1 - \rho_2/\rho_1)$, where ρ_1 and ρ_2 are the densities of the brass weight and gaseous medium respectively, and in the second to the electromagnetic force due to a current I_1 in the coil. Over a wide range of currents there is a strictly proportional relationship between the magnitudes of the current and the corresponding force, so that, with the voltage removed, the determination of the current I_2 required to balance the brass weight* will enable the force corresponding to any other current to be derived. In the case just considered, the force corresponding to I_2 is

$$mg(1 - \rho_2/\rho_1)\frac{I_1}{I_2}$$

The voltage is then calculated by means of eqn. (1) from the measured values of effective disc diameter, electrode spacing and force acting on the disc.

(5) INSTRUMENT CORRECTIONS AND ERRORS

The error in the voltage calculated from the equation

$$F_0 = \frac{\epsilon_0\epsilon_r\pi r^2 V^2}{2c^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

will depend on the extent to which the conditions assumed in the derivation of the equation have been fulfilled and on the errors of measurements of the parameters F_0, r and c. These questions are considered below and summarized in Table 1.

(5.1) Non-Uniformity of Field

Eqn. (1) gives the relation between the attractive force F on the disc due to an applied effective voltage V, on the assumption that the electrodes are perfectly flat and parallel and sufficiently extensive for the influence of the field disturbance at their edges to be ignored. To determine the extent to which the latter of these conditions is fulfilled the field distribution between the h.v. and earthed electrode systems was calculated for the maximum electrode spacing. The calculation, which was performed by relaxation methods, ignored the gaps between the guard plate and the disc and between the guard plate and the tank, and special attention was paid to the space between the flat parts of the electrodes. The graph of Fig. 5 has been derived from the results obtained, in order to show the relation between the voltage gradient at a point on the surface of the earthed electrode and the distance of the point from the centre. The gradient is expressed as a fraction of $E = V/c$, the value it would have if the electrodes were of infinite extent. Over the area covered by the disc, the relation between gradient E_s and distance s from the centre is closely given by the equation

$$E_s = E_0(1 - ks^3)$$

where E_0 = Gradient at centre

$k = 1\cdot 5\ (\text{metre})^{-3}$

Therefore the attractive force on the disc is

$$\frac{\epsilon_0\epsilon_r E_0^2}{2}\int_0^r 2\pi s(1 - ks^3)^2 ds$$

$$= \frac{\epsilon_0\epsilon_r\pi E_0^2}{2}\int_0^r 2(s - 2ks^4 + k^2s^7)ds$$

$$= \frac{\epsilon_0\epsilon_r\pi E_0^2}{2}(r^2 - \tfrac{4}{5}kr^5 + \tfrac{1}{4}k^2r^8)$$

$$= \frac{\epsilon_0\epsilon_r\pi r^2 E_0^2}{2}(1 - \tfrac{4}{5}kr^3 + \tfrac{1}{4}k^2r^6) \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The expression outside the brackets in eqn. (4) is equal to the attractive force if the gradient over the whole disc were equal to E_0; the expression within the brackets, after substituting the values of k and r, is $1\cdot 5$ parts in 10^4 less than unity. But E_0 is $0\cdot 4$ part in 10^4 less than E (see Fig. 5) so that the attractive force at an electrode spacing of 5 cm is less than it would be with electrodes of infinite extent by $1\cdot 5 + 0\cdot 8 = 2\cdot 3$ parts in 10^4. The derived voltage, being proportional to the square root of the force, will be in error by half this amount if the effect of the finite electrode size is ignored: this error will be still smaller with electrode spacings of less than 5 cm.

(5.2) Flatness of Electrodes

Although it is necessary for the disc to be coplanar with the guard plate to within about 1 micron to ensure voltage measurements accurate to 1 part in 10^4 (Section 2.3), it is not necessary for the electrodes to be flat to this degree. In general, for the above accuracy, the departures from a true plane of the surfaces of the disc, of the zone of the guard plate immediately surrounding it and of the h.v. electrode opposing it should be not more than one ten-thousandth of the electrode spacing; i.e. 5 microns at a spacing of 5 cm. Irregularities much greater than this can be tolerated in the peripheral parts of the electrodes; indeed, in the preparation of the surface of the guard plate, it was found necessary to take a light skim (approximately 50 microns) off the outer part so that the inner part could be worked to the requisite degree of flatness; the calculated effect of this on the field at the surface of the disc was quite negligible.

When tested with a knife-type straight-edge, the vital parts of the electrode surfaces were estimated to be flat to within $2\cdot 5$ microns, and the disc was adjudged to be coplanar with the guard plate to the same degree of accuracy; uncertainties in voltage measurement due to the combination of the uncertainties in these two factors are ± 3 and ± 5 parts in 10^4 at electrode spacings of 5 cm and 1 cm respectively.

(5.3) Parallelism of Electrodes

If the electrodes are not quite parallel, the attractive force on the disc is increased by $\beta^2/4$, where β is the difference between the maximum and mean separation of the disc from the h.v.

* The sign of the current I_2 will be opposite from that of I_1 since it has to balance a downward-acting force.

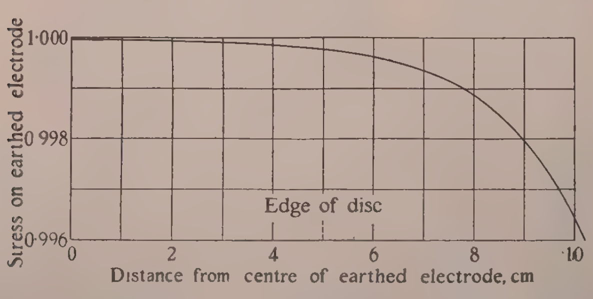


Fig. 5.—Relation between electric stress at a point on the earthed electrode (expressed as a fraction of the average stress between the electrodes) and distance of the point from the centre.

Electrode spacing—5 cm.

electrode, expressed as a fraction of the electrode spacing c (see Appendix 10). For a given angle between the planes of the electrodes, β will vary inversely as the electrode separation. For separations as small as 1 mm, however, the error in the attractive force is still negligible if this angle is not greater than 1′. With the adjustment available on the pillars supporting the guard plate it is possible to ensure that the electrodes are parallel to within 0·1′; errors due to lack of parallelism of the electrodes are therefore negligible.

(5.4) Dimensions of Electrodes

In calculating the voltage V from eqn. (1) the effective disc radius r and the effective electrode spacing c are directly concerned. The value of r is obtained from the following micrometer measurements, made at a temperature of 15° C:

Diameter of disc	9·9705 cm
Diameter of hole in guard plate	10·0279 cm

The mean value, 9·9992 cm, is only 4 parts in 10^6 less than the calculated effective diameter of the disc; hence $r = 4·9996$ cm. The errors due to uncertainty in the value of r and to any changes in this value due to changes of temperature are negligible.

The mean pitch of the screw threads on each of the three pillars supporting the earthed-electrode system was within 1 part in 10^4 of the nominal value of 1 mm. The electrode spacing c is therefore derived to this degree of accuracy from the differences between the reading of the micrometer scale during the voltage measurement and that corresponding to zero electrode spacing. The results of a typical determination of the effective zero are illustrated in Fig. 6, where the value of the

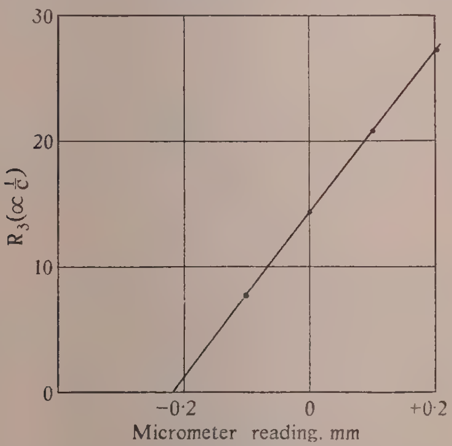


Fig. 6.—Relation between micrometer reading and reciprocal of capacitance C between h.v. electrode and attracted disc.

Contact zero at micrometer reading—0·200 mm.
Effective zero at micrometer reading—0·218 mm.

resistance arm R_3 of a bridge, which is proportional to the reciprocal of the capacitance between the disc and h.v. electrode, is plotted against the corresponding reading of the micrometer scale. The fact that a linear relationship between R_3 and micrometer reading, deduced from the formula for the capacitance between parallel plane electrodes, is closely obeyed down to an electrode separation of 0·1 mm provides good evidence of the perfection of form and of adjustment of the electrodes. The uncertainty in the extrapolated value of micrometer reading corresponding to $R_3 = 0$ is estimated to be $\pm 0·005$ mm; this leads to an uncertainty in voltage measurement of 1 part in 10^4 when $c = 5$ cm and 5 parts in 10^4 when $c = 1$ cm.

(5.5) Measurement of Attractive Force

The stiffness of the four phosphor-bronze strips supporting the disc is such that a load of 1 g causes a displacement of 8 microns; it thus just fulfils the condition, derived in Section 2.2, for a stable system at the highest designed stress between the main electrodes. The detector of displacement of the disc is sufficiently sensitive to measure the ratio of the capacitances on either side of the electrode K to 1 part in 10^3. Since the gaps between the moving and each of the fixed electrodes are about 1 mm, the detector is thus sensitive to a displacement of the moving system of 0·5 micron. The accuracy with which the attractive force on the disc can be balanced is therefore about double that derived in Section 2.3 for a displacement detector sensitive to 1 micron.

In addition to the above uncertainty of balance, there is an uncertainty in the measurement of the force (gravitational or electromagnetic) opposing the electrostatic pull on the disc. When the opposing force is gravitational, the uncertainty in its magnitude is negligible, but when it is electromagnetic there are errors in the measurement of the coil current and uncertainty in the relation between coil current and electromagnetic force to consider.

The coil current is measured either by a multi-range substandard moving-coil milliammeter readable over the upper part of its scale to an accuracy of 0·1%, or if higher accuracy is required, by a Campbell bifilar reflecting galvanometer used in conjunction with a semi-circular scale 4 m in length. Tests with loads on the disc ranging up to 0·5 kg have shown that there is a linear relation (to the accuracy with which balance could be adjusted) between load and coil current required for balance. When the galvanometer is used, it is therefore estimated that the electromagnetic force can be measured to an accuracy of 1 part in 10^3.

The chief sources of error considered in Sections 5.1–5.5 and

Table 1

Estimated Errors in Voltage Measurements

Source of error		Estimated effect			
		Electrode spacing = 1 cm		Electrode spacing = 5 cm	
		100 kV/cm	30 kV/cm	100 kV/cm	30 kV/cm
		parts in 10^4	parts in 10^4	parts in 10^4	parts in 10^4
Disc not coplanar with guard plate to better than 2·5 microns during initial setting-up		±5	±5	±3	±3
Effective zero of micrometer scale uncertain to 5 microns		±5	±5	±1	±1
Sensitivity of detector of displacement of disc limited to 0·5 micron		±1	±10	±1	±10
Error in measurement of coil current		±5	±5	±5	±5
Resultant standard error	Electro-magnetic balance	±9	±13	±6	±12
	Gravita-tional balance	±7	±12	±3	±10

the estimated extent to which they will affect voltage measurements at electrode spacings of 1 and 5 cm and at inter-electrode stresses of 100 kV/cm and 30 kV/cm are summarized in Table 1. The resultant standard errors are given in the last two lines of the Table for the cases in which the attractive force on the disc is balanced by an electromagnetic or a gravitational force; these errors are smallest when the spacing and electric stress between the electrodes are greatest.

(6) PERFORMANCE OF INSTRUMENT

(6.1) Flashover Voltage

After assembly, the instrument was connected to a 50-c/s supply through a protective resistance of 0·25 megohm, and the average value of the voltage required to cause an internal discharge was determined for a range of electrode spacings and gas pressures; no special precautions were taken to dry or filter the gas. The results obtained with the porcelain insulator are given in Fig. 7. With increasing electrode spacing the flashover

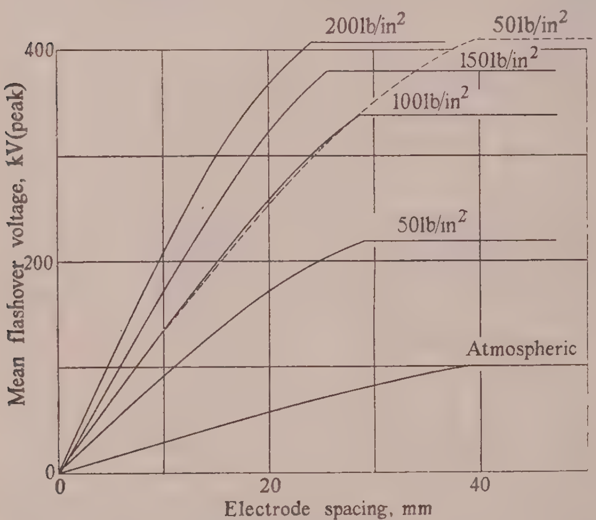


Fig. 7.—Relation between electrode spacing and flashover voltage using porcelain insulator.

——— Nitrogen.
— — — Nitrogen and Arcton 6.

The values on the curves refer to lb/in² gauge.

voltage at first increased and then remained constant, owing presumably, to the transference of the flashover path from the gap between the electrodes to another part of the system. Subsequent examination of the electrodes showed that many discharges had occurred between the stem of the h.v. electrode and the upper edge of the turret through which it passed; it is concluded that breakdown at this point limits the voltage which can be applied to the instrument when the porcelain insulator is used.

A voltage of 90% of the flashover value can be applied to the instrument for prolonged periods. Thus, with the porcelain insulator, voltages of up to 350 kV (peak) can be measured when nitrogen at a pressure of 200 lb/in² or a mixture of nitrogen and Arcton 6 at 50 lb/in² is used as a dielectric.

The curves of Fig. 7 show that, with nitrogen at a pressure of 150 lb/in² and over, a working stress of the value assumed in the design of the instrument, i.e. 140 kV/cm (Section 2.1), can be attained. However, in the range over which flashover occurs between the electrodes, the curves tend to flatten off with increasing electrode spacing, and it may be that, with the larger insulator, more care may be needed in filtration of the gas filling to enable this stress to be maintained without flashover up to the maximum electrode spacing.

(6.2) Comparison with Other Instruments

Simultaneous measurements of a 230-kV (r.m.s.), 50-c/s alternating voltage were made with the absolute voltmeter filled with a mixture of nitrogen and Arcton 6 at 50 lb/in² pressure and with a low-range electrostatic voltmeter with a scale 4 m in length used in conjunction with a capacitive voltage-divider. The voltage supply was unsteady to the order of 1 part in 10^3, but with one observer maintaining the balance of the attracted disc and another reading the adjacent scales of the electrostatic voltmeter and galvanometer, a comparison between the two instruments was obtained to a sufficiently high degree of accuracy. The results, given in Table 2, show agreement between the two methods of measurement to within the limits of the estimated errors given in Table 1.

Table 2

COMPARISON OF MEASUREMENTS MADE WITH ABSOLUTE VOLTMETER AND LOW-RANGE VOLTMETER USED IN CONJUNCTION WITH VOLTAGE DIVIDER

Relative permittivity of gas in absolute voltmeter, ϵ_r	1·0037
Density of gas in absolute voltmeter, ρ_2	0·0054 g/cm³
Density of 99·80 g brass weight, ρ_1	8·4 g/cm³
Gravitational force, g	981·0 dynes/g
Coil current required to balance 99·80 g weight, I_2	23·94 mA
Measured ratio of capacitance voltage-divider	2 020

Absolute voltmeter			Electrostatic voltmeter	
Effective electrode spacing (c)	Coil current required for balance	Calculated voltage	Instrument voltage (corrected)	Total voltage
mm	mA	kV(r.m.s.)	volts	kV(r.m.s.)
49·935	17·44	225·8	111·75	225·7
$42\cdot6_9$	*	226·1	111·85	225·9
39·935	27·28	225·8	111·75	225·7
34·935	35·80	226·3	112·07	226·4

* In this case the electrode spacing was adjusted until the attractive force on the disc was just balanced by the fixed weight.

As an example of the wide range of the instrument and of the effect of maladjustment of the disc, Table 3 gives the results of a comparison between the absolute voltmeter used at atmo-

Table 3

Dynamometer voltmeter reading	Absolute voltmeter				
	Micrometer reading	Effective electrode spacing	Net coil current	Displacement of disc above plane of guard plate†	Calculated voltage
volts (r.m.s.)	mm	mm	mA	microns	volts (r.m.s.)
784	0·200	0·431	2·83	0	784
		0·425	2·93	7	787
		0·417	3·05	15	788
		0·410	3·15	25	788
299	0·000	0·218	1·65	0	303
		0·211	1·78	7	305
		0·205	1·91	15	307
		0·198	2·09	25	310

† Calculated from the change in the setting of the audio-frequency bridge.

spheric pressure and a substandard dynamometer voltmeter. The maladjustment was effected by altering one of the resistance arms of the audio-frequency Schering bridge by a known amount and adjusting the current in the coil until balance at the new setting was obtained, first with no voltage and then with the voltage to be measured applied to the instrument. The difference between the two values of coil current was proportional to the attractive force on the disc and is referred to in the Table as the net coil current.

It will be seen from the Table that at small electrode spacings the method of determining the effective zero of the micrometer scale described in Section 5.4 compensates to a large extent for any failure to adjust the disc to be coplanar with the guard plate, and that an accuracy of about 1% can still be attained when measuring potentials of the order of a few hundred volts.

(7) CONCLUSIONS

The instrument here described, when fitted with the porcelain insulator and filled with nitrogen at a pressure of 150 lb/in^2 or a mixture of nitrogen and Arcton 6 at a pressure of 50 lb/in^2, is capable of measuring, to an accuracy of 0·1%, the effective value of sinusoidal voltages ranging from 30–250 kV (r.m.s.): at voltages below 30 kV the accuracy is reduced, but it is still of the order of 1% at 300 volts. The highest accuracy is obtained with the electrodes at large spacings and a high electric stress between them. The accuracy falls off rapidly as the stress between the electrodes—and hence the attractive force acting on the disc—is reduced. This could be compensated to some extent by increased sensitivity of the detector of displacement of the disc, but on the whole it is desirable that the attractive force should be comparable with the weight of the permanent load on the elastic supports of the disc system, otherwise imperfections in these supports may limit the accuracy with which the attractive force can be measured.

For routine measurements with an instrument of this type, the electrode spacing could be adjusted until the electrostatic force on the disc balanced a fixed weight so chosen that the micrometer scale could be read directly in kilovolts.

(8) ACKNOWLEDGMENTS

The work described above has been carried out as part of the research programme of the National Physical Laboratory, and the paper is published by permission of the Director of the Laboratory.

The instrument has been constructed in the Laboratory workshops, and the author desires to acknowledge the assistance in some of the details of design of the low-voltage electrode system and its associated measuring equipment given by members of the staff of the Metrology Division. The Mathematics Division was responsible for the calculation of the field distribution referred to in Section 5.1.

(9) REFERENCES

(1) Lord Kelvin: "Papers on Electrostatics and Magnetism," Second Edition, Section 360, p. 287 (Macmillan and Co., 1884).

(2) Brooks, H. B., Defandorf, F. M., and Silsbee, F. B.: "An Absolute Electrometer for the Measurement of High Alternating Voltages," *Journal of Research of the National Bureau of Standards*, 1938, **20**, p. 253.

(3) Palm, A.: "Vereinfachtes Absolutes Voltmeter in Pressgas," *Zeitschrift für technische Physik*, 1933, **14**, p. 390.

(4) Böcker, H.: "Ein Hochspannungsmesser für 600 kV," *Archiv für Elektrotechnik*, 1939, **33**, p. 801.

(5) Skilling, H. H., and Brenner, W. C.: "The Electrical Strength of Nitrogen and Freon under Pressure," *Transactions of the American I.E.E.*, 1942, **61**, p. 191.

(10) APPENDIX

Effect of Lack of Parallelism of the Electrodes on the Force acting on the Disc

Consider the surface of the disc (Fig. 8) to be divided into a number of elementary strips on each of which the electric stress is constant but, owing to lack of parallelism of the electrodes, is different from that on the other strips.

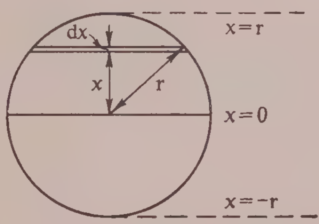


Fig. 8.—Diagram of disc to assist in the calculation of the effect of lack of parallelism of the electrodes on the attractive force.

Let the stress Ex on a strip distant x from the central strip and of width dx be $E(1 + \alpha x)$.

Attractive force on strip is

$$\epsilon_0\epsilon_r E^2(1 + \alpha x)^2\sqrt{(r^2 - x^2)}dx$$

Total force on disc is

$$\epsilon_0\epsilon_r E^2\int_{-r}^{r}(1 + \alpha x)^2\sqrt{(r^2 - x^2)}dx$$

$$= \epsilon_0\epsilon_r E^2\int_{-r}^{r}\left\{\sqrt{(r^2 - x^2)} + 2\alpha x\sqrt{(r^2 - x^2)} + \alpha^2x^2\sqrt{(r^2 - x^2)}\right\}dx$$

$$= \epsilon_0\epsilon_r E^2\left[\frac{x}{2}(r^2 - x^2)^{\frac{1}{2}} + \frac{r^2}{2}\arcsin\frac{x}{r} - \frac{2\alpha}{3}(r^2 - x^2)^{\frac{3}{2}} - \frac{\alpha^2x}{4}(r^2 - x^2)^{\frac{3}{2}} + \frac{\alpha^2r^2}{8}x\sqrt{(r^2 - x^2)} + \frac{\alpha^2r^4}{8}\arcsin\frac{x}{r}\right]_{-r}^{r}$$

$$= \epsilon_0\epsilon_r E^2\left(\frac{\pi r^2}{2} + \frac{\pi\alpha^2r^4}{8}\right)$$

$$= \frac{\epsilon_0\epsilon_r\pi r^2E^2}{2}(1 + \beta^2/4)$$

where $\beta = \alpha r$ is the amount by which the maximum (or minimum) stress on the disc differs from the mean stress.

[The discussion on the above paper will be found overleaf.]

DISCUSSION BEFORE THE MEASUREMENTS SECTION, 14TH DECEMBER, 1954

Professor E. Bradshaw: The elegant instrument described represents a considerable achievement, and the paper is, I believe, the first British account of an attracted-disc-type absolute voltmeter for this voltage range. The author has been conservative in the accuracy claimed and gives a satisfactory comparison between this instrument and the divider-operated low-range electrostatic voltmeter. Has not an opportunity been missed, however, to use as a comparison instrument the ellipsoid voltmeter developed by Bruce,* for which a somewhat higher accuracy is claimed? Although the author states that compressed gas was chosen as a medium of high electric strength because of the problem of leaks in vacuum systems, it would appear that if this is the only reason, modern vacuum technique should be capable of meeting this problem. There are certain philosophical and material objections to high gas pressure. If gas is used, the relative permittivity of the gas has to be measured. To this extent, the use of the word "absolute" appears to be slightly compromised. The author has discussed differential thermal expansion of the insulator and of the unavoidably long supporting stem of the h.v. electrode; has any uncertainty of the gap been observed due to dimensional instability of the insulator assembly under pressure?

The use of high working gradients results in satisfactorily small dimensions and large working forces. Are these forces capable of causing any distortion of the moving disc which is supported at the centre and free at the edges adjacent to the guard ring? The position of the moving disc is measured at the centre, and if even a small distortion of the disc occurs, lack of coplanarity with the guard ring may be significant.

The method employed for adjusting the separation of the fixed and moving electrodes is very ingenious, but I should like further information as to why the contact zero described in the paper differs from effective zero, and why effective zero is considered to be the right zero. Is this necessarily so?

In Table 2 I notice that the separation of about 42 mm is quoted to a much smaller degree of accuracy than any other case. I am not sure why that should be so.

The natural frequency of the moving system would, from the dimensions given, appear to be quite low. There is no question of the disc moving cyclically with the 100c/s electrostatic force. Has there been any trouble from oscillation of the moving system due to shock or floor vibration? There does not appear to be any provision of damping action against such oscillation.

Mr. R. Davis: When the decision was reached to construct an absolute voltmeter, the first question to be answered was whether to use compressed gas at about 14 atm or a high vacuum as the dielectric medium. Although both these media have a usable operating electric strength of about 100 kV/cm, the decision was in favour of compressed gas mainly because of the very practical advantage that in the event of a small leak the equipment could be connected to a high-pressure gas source and be usable; there is only one thing to be done in the event of a leak in a vacuum system. Other people have much more experience of vacuum technique than ourselves, and to them the problem of providing a vacuum of high electric strength would probably be faced with equanimity. Parts of the equipment are subjected to divergent fields, and while there is little precise information available about the flashover characteristics of electrode systems of different geometries in media of compressed gas and a high vacuum, it is well known that in equipment incorporating both compressed gases and a high vacuum, such as totally enclosed Van de Graaff generators, troubles are most frequently encountered on the vacuum side. The pressure chamber, it was thought, could also be used to study the dielectric properties of complex gases at high voltages as well as high stresses.

The absolute voltmeter measures direct voltages, but only the r.m.s. value of an alternating voltage; this latter fact must be regarded as a disadvantage when it is appreciated that, in general, high voltages are used in the laboratory to examine the electric strength or the flashover or breakdown of equipment. These characteristics depend on the peak applied voltage, so that when using the instrument for this purpose a further measurement is required. However, an instrument which will measure high direct voltages can be of more than local service. It can be used, for example, to calibrate resistors for use in high-voltage d.c. measurements.

Methods of measuring peak and r.m.s. voltages with the aid of a standard of capacitance are well known. In the discussion of papers on the calibration of sphere gaps,* one speaker deprecated the makeshift methods used for power-frequency voltage measurements and went on to say that "it is to be regretted that no effort has been made to develop a thoroughly screened condenser for 1MV." At a recent symposium at the N.P.L. on precision electrical measurements, an author described such a capacitor housed in a resin-bonded-paper tube and standing approximately 13 ft high. The limitation in the performance of this unit was imposed by external flashover, which was determined by the ambient humidity. Thus, at a relative humidity of 50% a test voltage of 1MV was sustained without external flashover; at 55% humidity external flashover occurred at 1MV; at 64% humidity external flashover started at 900 kV; and at 85% it was as low as 640 kV. Steps were being taken to avoid this limitation, possibly by the use of a special lacquer on the outside of the tube.

Mr. F. W. Waterton: When the author was starting work on his voltmeter we were also working on a similar project and were committed to the use of vacuum as the main insulating medium. Our experience has shown that the author's decision to use gas under pressure was sound, since the provision of mechanical insulating supports in the vacuum has been one of our major problems. We, of course, were aiming at a degree of accuracy one order higher than the author, as we already had a secondary standard for direct voltages which was accurate to $\pm 0\cdot 1\%$. The difficulties involved in improving the prospect of available accuracy by one order, in an instrument of this kind, bear no relation to the change in the order of accuracy.

Our instrument is generally similar in principle and electrode arrangement to the author's, except that we arranged to measure the force on the moving disc directly by means of a chemical balance without any intermediate steps and without the use of springs to control the disc position. It always seemed to us that this was by far the best method, as it did not involve the transfer of the force quantity from one medium to another, with the attendant possibility of loss of accuracy and the necessity for precautions to avoid the effects of temperature changes, spring fatigue, etc., when making a measurement, or during adjustments to the zero position of the disc.

The use of a chemical balance to support the moving disc in a vacuum is the method which involves the minimum number of possibilities of error and the minimum number of corrections. Against this advantage can be cited the greater ease and latitude in technique and in the choice of materials which follows on the use of gas at pressure as opposed to vacuum for the main insulating medium. Compared with any other medium, the use of

* BRUCE, F. M.: "The Design of an Ellipsoid Voltmeter for the Precision Measurement of High Alternating Voltages," *Journal I.E.E.*, 1947, **94**, Part II, p. 129.

* *Journal I.E.E.*, 1938, **82**, p. 645.

a vacuum allows a higher stress to be used between the electrodes and so enables a smaller disc to give the same force.

It would appear from Section 7 of the paper that the author's experience with his instrument would now lead him to subscribe to our views concerning the desirability, by the direct method, of weighing the pull on the disc, particularly for an instrument designed, as in our case, for the routine measurement of voltage in the laboratory and for the calibration of secondary standards of voltage.

I am very interested in the performance figures and design of the insulators shown in Figs. 1(*a*) and 1(*b*). From experiments and measurements carried out in a 700kV transformer insulated with CCl_2F_2 at 60 lb/in^2, I feel that the design shown in Fig. 1(*a*) would give trouble at a stress of 300, and at the most 350, volts/mil in the gas.

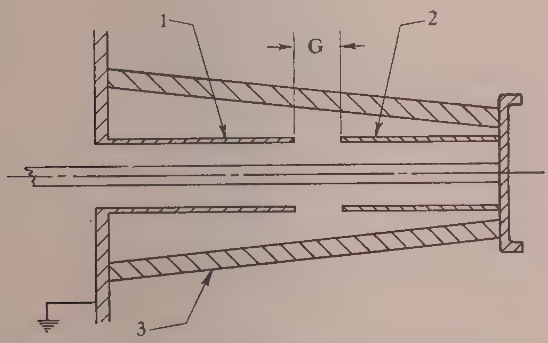


Fig. A.—Use of concentric cylinders to increase sparkover voltage of bushing.

The position of highest stress occurs at the point where the central conductor passes through the tank wall, and if the illustrations are to scale, the stress at this point would appear to be 550 volts/mil at the rated voltage. If sulphur hexafluoride were used these figures would be improved by about 15–20%, but if a stress of 550 volts/mil is present, a pressure higher than 60 lb/in^2 of sulphur hexafluoride will almost certainly be necessary.

The performance of both of these insulators could be improved if the radial stresses were limited to those suggested in the preceding paragraph by the introduction of a pair of concentric cylinders 1 and 2, such as are shown in Fig. A. If the gap G between the ends of such a pair of cylinders is adjusted experimentally to the optimum value, improvements of 75% or more in the sparkover value of the insulator 3 have been obtained on several different designs of bushing compared with an otherwise identical arrangement in which the cylinders were omitted.

In the paper the author shows the moving disc supported by leaf springs. Was this arrangement, in fact, used in the final instrument? I am rather intrigued by this, as I have attempted to use a similar system in a vacuum-enclosed instrument and found that the springs introduced considerable variations in the zero of the instrument, mainly due to changes in temperature, and these variations were of sufficient magnitude to cause the design to be abandoned.

Mr. F. S. Edwards: When an instrument is devised for very-high-precision measurements, its inventor always examines so far as he can all the sources of inaccuracy and makes allowances for them; but he can never be sure that something has not been overlooked. One or two possible errors, not referred to by the author, are, in fact, mentioned in the discussion, and there may be others.

The most convincing test is to compare the instrument with another of different construction and of the same or greater claimed accuracy. If the two instruments differ by more than the sum of their admitted possible errors, then at least one of them is not as accurate as it is said to be. If they agree there is a probability that both are correct. Mr. Bowdler has performed this test with gratifying results, but I should like to know whether the comparison was made at any intermediate voltages between a few hundred volts and 225kV, and whether the results were closely repeatable from day to day.

The capacitance of the voltage divider referred to in Section 6.2 could be measured with adequate accuracy at a low voltage, but what evidence is there that it remained unchanged at 225kV?

THE AUTHOR'S REPLY TO THE ABOVE DISCUSSION

Mr. G. W. Bowdler (*in reply*): The contributors to the discussion have shown a lively appreciation of the difficulties encountered in the accurate measurement of high voltages with an absolute instrument. I agree with Professor Bradshaw and Mr. Waterton that a hard vacuum is the ideal dielectric medium for such an instrument, but all available experience suggested that a vacuum capable of withstanding the voltages we had in mind would be very difficult to realize. The relative permittivites of nitrogen and other gases used in the instrument differ very little from unity, and in general the published values are sufficiently accurate for our purpose.

Professor Bradshaw rightly observes that high dimensional stability of the insulator is required. This has been so for the porcelain insulator shown in Fig. 1(*b*); in fact, the change in length of this insulator due to the pressures used so far has been too small to detect with certainty. This fact, together with the design of the support for the earthed electrode shown in Fig. 2, has resulted in an electrode spacing which is independent of gas pressure. Preliminary tests with the insulator shown in Fig. 1(*a*) indicate that its extension with pressure is by no means negligible; this may require a more frequent check of the contact zero of the electrode-spacing scale when this insulator is used. I am indebted to Mr. Waterton for his suggestion for increasing the voltage which a bushing insulator will withstand. The only reasons why the contact zero should differ from the effective zero are either that the disc is not coplanar with the guard plate or that the high-voltage and guard plates are not flat and/or parallel to each over their whole extent: the second of these alternatives is considered to be the more likely, and for this reason the effective zero was used to obtain the electrode spacing.

The attracted disc is sufficiently thick—even without its stiffening ribs—to resist appreciable deformation by the electrostatic force acting on it. The addition of the ribs probably puts the frequency of any natural vibrations of the disc well above 100c/s. Free or forced vibrations of the disc system on its elastic support appear to be effectively suppressed by eddy currents induced in the coil former.

The four phosphor-bronze strips by which the disc is supported have behaved very satisfactorily; their strain due to the weight they carry undergoes a secular change (probably as a result of temperature changes) of about 1 part in 1 000. This means that the unloaded disc cannot be assumed always to be coplanar with the guard plate: the force necessary to bring about this condition must be allowed for in voltage measurements.

I agree with Mr. Waterton that balance of the electrostatic force by a weight is the most satisfactory way of measuring it. With the single weight available in the present instrument this involves variation of the electrode spacing. This has been done

in the second line of Table 2, where the entry in the first column has been quoted to an accuracy corresponding to the sensitivity of the detector of displacement of the disc. The direct-reading feature mentioned in Section 7 of the paper is attractive; it remains to be seen to what extent the accurate screw-threads supporting the earthed electrode will wear under constant use.

The instrument has been used at many voltages intermediate between those quoted in Tables 2 and 3, and when the larger insulator is fitted, measurements will be made at as high a voltage as possible. The low-range electrostatic voltmeter used in conjunction with a capacitive voltage-divider as a comparison standard is a well-tried instrument which has all the accuracy required and which was readily available. The high-voltage arm of the divider consisted of a capacitor with compressed-gas dielectric and concentric cylindrical electrodes. Capacitors of this type but of different voltage ratings, when compared on a Schering bridge, are found to have no difference in power factor and a constant ratio of their capacitances at voltages up to that causing flashover. There is no reason, therefore, to suspect that the ratio of the capacitive voltage-divider is not constant.

DISCUSSION ON "EXPERIMENTAL EQUIPMENT AND TECHNIQUES FOR A STUDY OF MILLIMETRE-WAVE PROPAGATION"*

Dr. J. Collard (*communicated*): In Section 3.3.2 the authors mention what they call "metal-film bolometers" and say that these would repay further attention. It may save them a considerable amount of work to know that an instrument of this type was developed during the 1939–45 War, by the organization with which I am associated. It was first described in British Patent Specification No. 572881, and later in a paper.†

As originally developed for use at 3 cm, the resistive film formed part of one of the smaller walls of a rectangular guide and so absorbed only a small part of the power passing down the guide, thus allowing the remainder to be used for other purposes. However, it was pointed out in the paper referred to above that, if the power to be measured was small, it would be better to put the film across the guide so that it would act as a termination to the guide and thus absorb all the power. The resistive film consisted of a sputtered platinum film on glass.

In 1949, when a power-measuring device for a wavelength of 8 mm was required, an enthrakometer using a film across the guide was developed, and several of these instruments have been in successful use for many years. The authors include in their References one [No. (11)] which refers to this 8 mm enthrakometer but, apparently, they have not realized that the enthrakometer was the instrument which they describe as a metal-film bolometer. It was pointed out in that Reference that the enthrakometer is free from the serious errors to which the thermistor is subject at millimetric wavelengths.

Mr. W. E. Willshaw, Dr. H. R. L. Lamont and **Mr. E. M. Hickin** (*in reply*): We are grateful to Dr. Collard for drawing attention to his published work on the enthrakometer and for emphasizing the importance for millimetre waves of measurement of power by absorption in thin films, rather than in a resistive wire.

* WILLSHAW, W. E., LAMONT, H. R. L., and HICKIN, E. M.: Paper No. 1761 R, January, 1955 (see **102B**, p. 99).

† COLLARD, J.: "The Enthrakometer, an Instrument for the Measurement of Power in Rectangular Guides," *Journal I.E.E.*, 1946, **93**, Part IIIA, p. 1399.

DISCUSSION ON "AN INVESTIGATION OF THE CHARACTERISTICS OF CYLINDRICAL SURFACE WAVES"* AND "SURFACE WAVES"†

NORTH-EASTERN RADIO AND MEASUREMENTS GROUP, AT NEWCASTLE UPON TYNE, 6TH DECEMBER, 1954

Mr. J. Bilbrough: Has any work been done in grading the dielectric constant of the transmission-wire covering, and would it help to use a laminar covering material with the higher dielectric constant on the outside, with the possibility of preventing radiation caused by water droplets?

Could the direct radiation between transmitting and receiving horns be reduced by using a resonant launching system, which might consist of a circular metal plate having a large hole in the centre so spaced from the transmitting horn as to cancel direct radiation?

Prof. H. E. M. Barlow, Dr. A. E. Karbowiak and **Dr. A. L. Cullen** (*in reply*): It is quite possible that radiation from raindrops could be minimized in the way suggested by Mr. Bilbrough. The outer dielectric must be so chosen that the losses associated with it are less than the loss due to scattering of power by raindrops if the method is to be effective in reducing overall loss. Perhaps expanded polythene could be used.

The use of a resonant launching system might very well enhance mode purity, but an improvement in overall launching efficiency would only be achieved if the losses associated with the resonance were less than the radiation loss of the original system.

So far we have undertaken no experimental work on either of these topics.

* BARLOW, H. E. M., and KARBOWIAK, A. E.: Paper No. 1462 R, April, 1953 (see **100**, Part III, p. 321).

† BARLOW, H. E. M., and CULLEN, A. L.: Paper No. 1482 R, April, 1953 (see **100**, Part III, p. 329).

621.372.8

Paper No. 1786 R
May 1955

AN EXPERIMENTAL INVESTIGATION OF AXIAL CYLINDRICAL SURFACE WAVES SUPPORTED BY CAPACITIVE SURFACES

By Professor H. E. M. BARLOW, Ph.D., B.Sc.(Eng.), Member, and A. E. KARBOWIAK, Ph.D.

(*The paper was first received 9th August, and in revised form 3rd November*, 1954.)

SUMMARY

This paper records experimental work designed to investigate the use of a guide with a capacitive surface impedance for the support of an axial cylindrical surface wave. Previously only inductive surfaces had been employed for the purpose.

A solid dielectric rod, made of Perspex, was used as the guide, and its diameter was carefully adjusted to make the surface impedance come within the particular range of interest.

The experiments show that there is no doubt at all about the physical reality of an axial cylindrical surface wave supported by a capacitive surface and having a phase velocity *exceeding* that of light. Moreover, it has been established that, within the limits of experimental error, such waves vanish completely when the threshold condition necessary for their support in the capacitive region is approached, exactly as predicted by theory.

LIST OF SYMBOLS

r, θ, x = Cylindrical co-ordinates (Fig. 1).
A, B = Constants.
$H_n^{(1)}$ = Hankel function of the first kind and order n.
J_n = Bessel function of order n.
$\gamma = \alpha + j\beta$ = Axial propagation coefficient.
α = Axial attenuation coefficient.
β = Axial phase coefficient.
$u = a - jb$ = Radial propagation coefficient.
a = Radial decay coefficient.
b = Radial phase coefficient.
$K_0^2 = -\omega^2\mu_0\epsilon_0$ = Free-space wave number.
$K_1^2 = -\omega^2\mu_0\epsilon_1$ = Wave number.
$(K_1')^2 = -\omega^2\mu_0\epsilon_1'$ = Complex wave number.
λ_0 = Free-space wavelength.
$\mu_0\epsilon_0$ = Permeability and permittivity of free space, respectively.
$\mu_1\epsilon_1$ = Permeability and permittivity of the dielectric (Perspex).
$\epsilon_1' = \epsilon_1(1 - j\tan\delta)$ = Complex permittivity.
ϵ_r = Relative permittivity.
$\tan\delta$ = Loss tangent of the dielectric.
h_1^2 = Separation constant.
$Z_s = |Z_s|\underline{/\phi_s}$ = Surface impedance [cf. eqn. (3)].
η and ξ = Quantities defined by eqn. (5).
$z_s = r_s + jx_s$ = Quantity proportional to the normalized surface impedance [defined by eqns. (10) and (11)].
τ = Quantity defined by eqn. (8).
X = Quantity defined in Section 3.3.1.

(1) INTRODUCTION

The single-wire transmission line supporting a cylindrical surface wave is well known. Such waves, in travelling along the outside of a conductor of circular cross-section, do not radiate, so that, neglecting losses in the guide, the energy is propagated exclusively along its surface. The Zenneck wave, which behaves in a similar way over a flat surface, has been put forward as a particular case of the cylindrical surface wave when the guide is of infinite radius. There is, however, one important distinguishing characteristic between the usual forms of these waves. Thus the cylindrical surface wave has always been characterized by a phase velocity which is less than the free-space value, whereas the Zenneck wave exhibits a phase velocity exceeding that value.

Owing to experimental difficulties, the problem of following the transitional changes from one form of this surface wave to the other has so far not been satisfactorily resolved, and the difference between the phase velocities has tended to suggest that there is perhaps something of basic importance in its explanation.

In a recent paper[2] it was shown theoretically that a cylindrical surface wave of the $E_{0\times}$ type should be capable of support by a guide having a capacitive surface impedance within defined limits, and that in those circumstances the phase velocity of the wave along the interface with the surrounding medium should exceed the corresponding free-space value. Previously it had been assumed that only guides with a positive surface reactance were capable of providing the service required, and the present work was undertaken in order to ascertain experimentally the threshold conditions for the support of such waves, which are apparently more closely akin to the Zenneck wave.

In pursuing this objective, several forms of guide of circular cross-section capable of providing a surface impedance with a capacitive as well as a resistive component were considered. Thus the case was examined theoretically of a metal rod surrounded by a stratified dielectric consisting of a thin annular layer next to the conductor having a permittivity less than that of the outer medium. Alternatively, consideration was given to a corrugated form of guide made up of a series of metal discs and dielectric spacers mounted alternately side by side along the length. These arrangements, whilst theoretically possible, presented difficult practical problems, particularly in bringing the surface impedance within the range of values required. It was decided, therefore, to use the simpler construction of a solid dielectric rod made of Perspex and having a diameter which gave the required surface impedance, taking into account the losses in the material.

This arrangement had the advantage that the guide was easily machined, and its diameter could be reduced in stages to enable the transition from an inductive to a capacitive surface impedance to be followed. It was necessary to know accurately the permittivity and loss tangent of the Perspex rod employed. Measurements of these quantities were made on samples of material cut from the actual guide.

The practical aspect of launching a cylindrical surface wave of the $E_{0\times}$ type has already been discussed in detail by the authors[1] with reference to metallic guides, and in that paper it was pointed out that, for accurate measurements on the field distribution of such a wave, recourse to a surface-wave resonator was most desirable, since this could be used to enhance the purity of the mode.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Dr. Barlow is Professor of Electrical Engineering at University College, London.
Dr. Karbowiak is with Standard Telecommunication Laboratories, Ltd., and was formerly at University College, London.

In the present investigation the same technique was employed, and consequently only a brief outline of the method will be given.

(2) FUNDAMENTALS OF THE THEORY

It will be assumed that the media involved are homogeneous and isotropic. The rationalized M.K.S. system of units is used throughout, and the factor $\varepsilon^{(j\omega t - \gamma x)}$ is omitted for convenience.

(2.1) Equations pertaining to the Axial Cylindrical Surface Wave (E_{0x} Mode)

The E_{0x} mode, which exists in the space outside a guide of circular cross-section (Fig. 1) having a radius s (i.e. for $r > s$), is characterized by the three field components:[1]

$$\left.\begin{aligned} E_x &= A\mathrm{H}_0^{(1)}(jur) \\ E_r &= A\frac{\gamma}{ju}\mathrm{H}_1^{(1)}(jur) \\ H_\theta &= A\frac{\omega\epsilon_0}{u}\mathrm{H}_1^{(1)}(jur) = -\frac{A}{Z_0}\frac{jK_0}{u}\mathrm{H}_1^{(1)}(jur) \end{aligned}\right\} \quad (1)$$

where
$$\gamma^2 + u^2 = K_0^2 = -\omega^2\mu_0\epsilon_0 \quad (2)$$

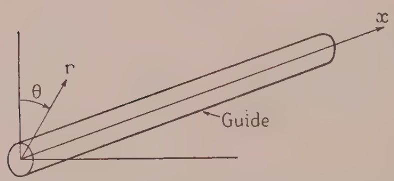


Fig. 1.—Cylindrical surface waveguide and co-ordinate system employed.

The ratio E_x/H_θ at the surface of the guide, where $r = s$, is termed the guide surface impedance, Z_s, and is given by

$$Z_s = \left.\frac{E_x}{H_\theta}\right]_{r=s} = \frac{u}{\omega\epsilon_0}\frac{\mathrm{H}_0^{(1)}(jus)}{\mathrm{H}_1^{(1)}(jus)} \quad (3)$$

When the argument of the Hankel functions in eqn. (3) is small (say less than 0·05), a good approximation to the solution of this equation can be obtained from

$$\eta = \xi \log_\varepsilon \xi \quad (4)$$

where
$$\left.\begin{aligned} \eta &= 1\cdot584 K_0 s\left|\frac{Z_s}{Z_0}\right| \angle\phi_s \\ \xi &= (0\cdot89us)^2 = |\xi| \angle\phi_\xi \end{aligned}\right\} \quad (5)$$
and
$$Z_s = |Z_s| \angle\phi_s$$

When the surface impedance, Z_s, of the guide is small and uniformly distributed over the surface the quantity u ($|u| \ll |K_0|$) is obtained through the solution of eqns. (4) and (5), and the propagation coefficient γ follows from eqn. (2), namely

$$\gamma = \alpha + j\beta \quad (6)$$

where
$$\left.\begin{aligned} \alpha &= \frac{\lambda_0}{2\pi}(ab) \\ \beta &= \frac{2\pi}{\lambda_0}(1 + \tfrac{1}{2}\tau) \end{aligned}\right\} \quad (7)$$

and
$$\tau = \left(\frac{\lambda_0}{2\pi}\right)^2(a^2 - b^2) + \left(\frac{\lambda_0}{2\pi}\right)^4(a^2b^2) \quad (8)$$

Sometimes, however, it is not legitimate to replace the Hankel functions, occurring in eqn. (3), by their small argument approximations to arrive at eqn. (4). This occurs, for example, in problems connected with predominantly reactive surfaces, and in that case the quantity a (here $a \gg b$ and hence $|u| \simeq a$) is obtained by graphical solution, as shown in Fig. 2, of

$$jX_s = \frac{a}{\omega\epsilon_0}\frac{\mathrm{H}_0^{(1)}(jas)}{\mathrm{H}_1^{(1)}(jas)} \quad (9)$$

Turning attention to eqns. (4) and (5), we observe that for certain ranges of the impedance angle, ϕ_s, the solution of eqn. (4)

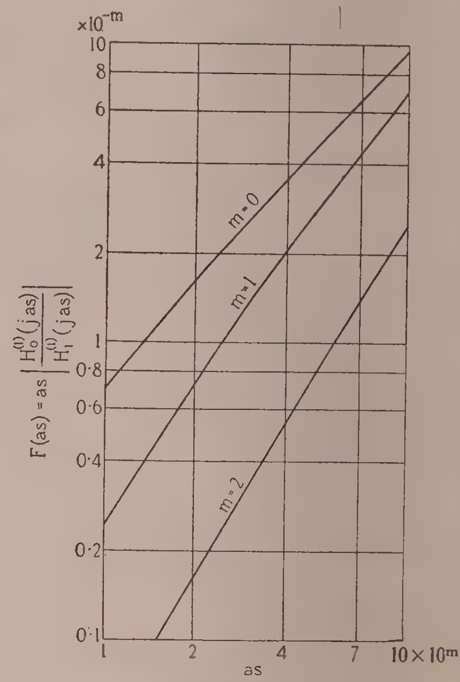


Fig. 2.—Curve of $as\left|\frac{\mathrm{H}_0^{(1)}(jas)}{\mathrm{H}_1^{(1)}(jas)}\right|$ as a function of as.
m = Scale index.

is quite straightforward, but for the general case of any value of ϕ_s and $|Z_s| \ll Z_0$, the following approach has been developed.

The quantity η can, with advantage, be put into the form

$$\eta = x_s - jr_s \quad (10)$$

where r_s and x_s are both dimensionless quantities proportional, respectively, to the normalized surface resistance R_s/Z_0 and the normalized surface reactance X_s/Z_0.

Thus we have

$$\left.\begin{aligned} r_s &= 1\cdot584\left(\frac{2\pi s}{\lambda_0}\right)\left(\frac{R_s}{Z_0}\right) \\ x_s &= 1\cdot584\left(\frac{2\pi s}{\lambda_0}\right)\left(\frac{X_s}{Z_0}\right) \end{aligned}\right\} \quad (11)$$

Eqn. (4) can then be split into its real and imaginary parts, giving

$$\left.\begin{aligned} -r_s &= |\xi|(\sin\phi_\xi \log_\varepsilon|\xi| + \phi_\xi\cos\phi_\xi) \\ +x_s &= |\xi|(\cos\phi_\xi \log_\varepsilon|\xi| - \phi_\xi\sin\phi_\xi) \end{aligned}\right\} \quad (12)$$

It will be observed that for given values of s and u, the corresponding values of r_s and x_s can be calculated from eqn. (12),

so that, with the help of eqn. (11), R_s and X_s are obtained. Although the reverse procedure is considerably more laborious, eqn. (4) or its equivalent, eqn. (12), can be solved with the help of an appropriate chart. Fig. 3 represents contours of constant $|\xi|$ and constant ϕ_ξ plotted in the plane of $z_s = r_s + jx_s$.

Since a surface wave can only exist for finite positive values of the components a and b of the radial propagation coefficient $u = a - jb$, hence $0 \geqslant \phi_u > -90°$, where $u = |u|\varepsilon^{j\phi_u}$. It follows from eqn. (5) that $\phi_\xi = 2\phi_u$ can take values from 0 to $-180°$. It is therefore apparent from the chart in Fig. 3 that a

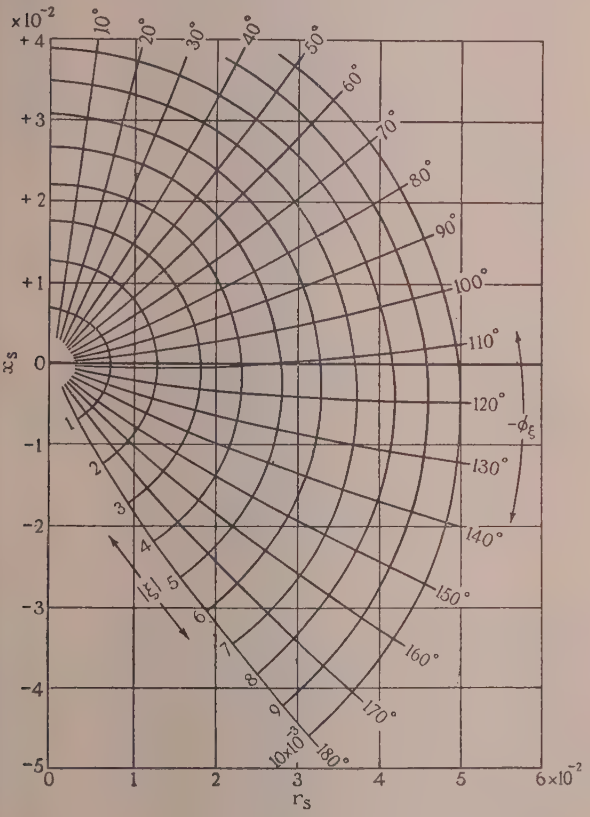


Fig. 3.—Chart of the complex quantity ξ in terms of r_s and x_s.

cylindrical surface wave could, at least theoretically, be supported by a purely resistive surface ($\phi_s = 0$) or even by a capacitive surface ($\phi_s < 0$). A surface wave supported by a capacitive surface would, as can be shown from eqn. (8), travel with a phase velocity exceeding that of the corresponding TEM wave in free space (this is a feature of the Zenneck wave). It is, however, to be noted that to satisfy the required conditions in the capacitive region the surface impedance must possess a substantial resistive component, and this must be such that $\phi_s = \arctan x_s/r_s$ does not exceed a certain value Φ_s, for which $\phi_\xi = -180°$, given by

$$\Phi_s = \arctan\left(\frac{\log_\varepsilon |\xi|}{\pi}\right) \quad . \quad . \quad . \quad . \quad (13)$$

The quantity τ can, if required, be expressed in terms of the angle ϕ_ξ as follows:

$$\tau = \left(\frac{|u|\lambda_0}{2\pi}\right)^2 \cos\phi_\xi + \frac{1}{4}\left(\frac{|u|\lambda_0}{2\pi}\right)^4 \sin^2\phi_\xi \quad . \quad . \quad (14)$$

In our present applications $|u| \ll 2\pi/\lambda_0$ ($|u|\lambda_0/2\pi$ is invariably less than $0 \cdot 1$), so that unless ϕ_ξ is close to 90°, we may assume

$$\tau = \left(\frac{|u|\lambda_0}{2\pi}\right)^2 \cos\phi_\xi \quad . \quad . \quad . \quad . \quad . \quad (15)$$

From this it is obvious that if ϕ_ξ is greater than 90° so that τ is negative, from eqn. (7) $\beta < 2\pi/\lambda_0$ and the wave will travel with a velocity greater than that of light.

(2.2) Theory of a Lossy Dielectric-Rod Waveguide (supporting the E_{0x} Mode)

Imagine a guide consisting of a dielectric rod (constants μ_1, ϵ_1 and σ_1) of uniform circular cross-section and radius s. If the field outside the guide is an E_{0x} wave, the axial component of the electric field inside the guide is given by

$$E_x = BJ_0(h_1 r) \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

where

$$h_1^2 = \gamma^2 - (K_1')^2 \quad . \quad . \quad . \quad . \quad . \quad (17)$$

and the circumferential component of the magnetic field is

$$H_\theta = B\frac{j\omega\epsilon_1'}{h_1}J_1(h_1 r) \quad . \quad . \quad . \quad . \quad . \quad (18)$$

Thus the guide surface impedance is

$$Z_s = \left.\frac{E_x}{H_\theta}\right]_{r=s} = \frac{h_1}{j\omega\epsilon_1'}\frac{J_0(h_1 s)}{J_1(h_1 s)} \quad . \quad . \quad . \quad . \quad (19)$$

The system of eqns. (2), (17), (3) and (19) could in principle be solved for the axial propagation coefficient, γ, and the radial propagation coefficient, u, of the surface-wave field. The exact solution is, however, impracticable, and therefore the following approach was adopted.

Consider for the moment a loss-free dielectric rod ($\epsilon_1' = \epsilon_1$)—in which case the surface impedance will be a pure imaginary quantity ($Z_s = jX_{s0}$). In particular, if $h_1 s$ is just greater than $2 \cdot 4048$ [the first root of $J_0(x) = 0$] the surface impedance is inductive, and when it is less than this value the surface impedance is capacitive. Moreover, if values of h_1 and s are chosen such that $h_1 s$ is very close to $2 \cdot 4048$, $|Z_s|$ is very small and so is $|u|$. From eqns. (2) and (17) we get for this case

$$\left.\begin{aligned} h_1 \simeq h_0 &= \omega\sqrt{[\mu_0(\epsilon_1 - \epsilon_0)]} \\ \text{or} \qquad h_0 &= \frac{2\pi}{\lambda_0}\sqrt{(\epsilon_r - 1)} \end{aligned}\right\} \quad . \quad . \quad . \quad (20)$$

The corresponding surface impedance is a pure reactance, and is given by

$$\left.\begin{aligned} Z_s \simeq jX_{s0} &= \frac{h_0}{j\omega\epsilon_1}\frac{J_0(h_0 s)}{J_1(h_0 s)} \\ \text{or} \qquad jX_{s0} &= jZ_0\frac{\sqrt{(\epsilon_r - 1)}}{\epsilon_r}\left[-\frac{J_0(h_0 s)}{J_1(h_0 s)}\right] \end{aligned}\right\} \quad . \quad . \quad . \quad (21)$$

This is an approximate expression in so far as eqn. (20) is an approximation, and in general from eqns. (2) and (17) we have

$$h_1^2 = h_0^2 - u^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

or $h_1 = h_0 + \Delta h$, where $\Delta h \simeq -u^2/2h_0$, neglecting higher-order terms in the expansion.

From eqn. (19) we then have from Taylor's theorem, using the same degree of approximation,

$$Z_s \simeq \frac{h_0}{j\omega\epsilon_1}\left[\frac{J_0(h_0 s)}{J_1(h_0 s)} - s\Delta h\right] = jX_{s0} + \Delta_1(Z_s)$$

where

$$\Delta_1(Z_s) = -s\Delta h\left(\frac{h_0}{j\omega\epsilon_1}\right) = \frac{su^2}{j2\omega\epsilon_1} = -j\frac{Z_0}{2}\frac{\sqrt{(\epsilon_r - 1)}}{\epsilon_r}\frac{u^2}{h_0}s \quad . \; (23)$$

If the dielectric is lossy ϵ_1 must be replaced by ϵ_1', and if the loss angle of the dielectric is small (as with Perspex), from eqns. (17) and (2) we get

$$h_1^2 = K_0^2 - u^2 - (K_1')^2 = K_0^2 - u^2 - K_1^2(1 - j\tan\delta)$$

and we now find that to satisfy

$$h_1 = h_0 + \Delta h$$

we must have

$$\Delta h = -\frac{u^2}{2h_0} + j\frac{K_1^2\tan\delta}{2h_0} \quad . \; . \; . \; . \; (24)$$

This leads to the following expression for Z_s:

$$Z_s = \left[1 - \frac{u^2}{2h_0^2} + \frac{j}{2}\left(\frac{\epsilon_r - 2}{\epsilon_r - 1}\right)\tan\delta\right]\left[jX_{s_0} + \Delta_1(Z_s) + \Delta_2(Z_s)\right] \quad . \; . \; . \; . \; (25)$$

$$\left.\begin{aligned} \text{where} \quad & \Delta_1(Z_s) = -j\frac{Z_0}{2}\frac{\sqrt{(\epsilon_r - 1)}}{\epsilon_r}\frac{u^2}{h_0}s \\ \text{or} \quad & \Delta_1(Z_s) = j\frac{Z_0}{2}\left(\frac{u}{K_1}\right)^2\left(\frac{2\pi s}{\lambda_0}\right) \\ \text{and} \quad & \Delta_2(Z_s) = \frac{Z_0}{2}\left(\frac{2\pi s}{\lambda_0}\right)\tan\delta \end{aligned}\right\} \quad . \; . \; . \; (26)$$

The multiplying factor in the first square brackets of eqn. (25) can usually be taken to equal unity, and if the decay coefficient u is small, $\Delta_1(Z_s)$ can be neglected. Consequently, under such conditions, knowing the dimensions of the guide and the constants of the dielectric (of which the guide is made), the surface impedance is calculable from eqns. (25) and (26). Values of surface impedance so obtained are then substituted in eqn. (5), giving corresponding values of η, ξ and thence u. Fig. 3 is plotted from eqn. (12) for a range of values of ξ, and greatly assists the solution of these equations. If the coefficient u is not sufficiently small to justify the neglect of $\Delta_1(Z_s)$, the solution of the problem is arrived at by successive approximations.

It will be observed that the resistive component $\Delta_2(Z_s)$ of the surface impedance arises as a result of dielectric loss, while the surface reactance is determined by the diameter of the dielectric rod in relation to frequency and the permittivity of the dielectric concerned.

(3) EXPERIMENTAL WORK

In attempting to launch a cylindrical surface wave over a capacitive surface and to examine under threshold conditions the distribution of the field, all the experiments were made in the 3 cm waveband.

The principal instrument employed in the measurements was the surface-wave resonator (Fig. 4), consisting of a length of guide made from a rod of Perspex short-circuited at both ends by large metal plates (about 4 ft × 4 ft) and mounted at right-angles to the guide. The resonator was excited by a small annular opening over the guide at the input and formed by the open end of a coaxial-line feed, as shown in Fig. 4.

The purity of a surface wave supported by the resonator was estimated by examining the relevant radial decay curve of the surface-wave field. Such decay curves, which represent the magnetic field strength as a function of radial distance from the centre of the guide (Fig. 5), were obtained experimentally by means of a radial-field-measuring instrument mounted on the terminating plate of the resonator. A quantitative examination of a particular decay curve was carried out in each case with the help of the corresponding identifying line (see Section 3.2.3). In each case the resonant frequency of the resonator was measured (to within about 1 part in 10^4) using a cavity wavemeter, while the guide wavelength was calculated (to within 5 parts in 10^4) from a knowledge of the total length of the guide and the number of wavelengths within it.

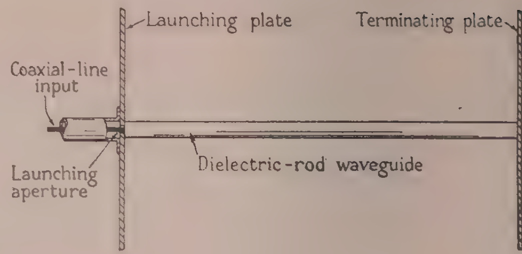


Fig. 4.—Surface-wave resonator with dielectric-rod waveguide.

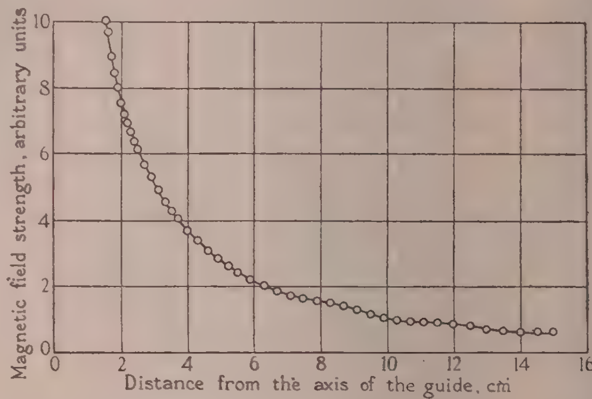


Fig. 5.—Magnetic-field distribution curve for guide No. 2, and giving point A in Fig. 13.

Guide length = 116·90 cm.
Guide radius = 0·97815 cm.
λ_0 = 3·2060 cm.

(3.1) The Guides

A length of 1 in-diameter unplasticized Perspex rod was used for preparing the guides, and they were made in two stages as follows.

Guide No. 1 was obtained by machining the Perspex rod to 0·817 ± 0·0005 in diameter corresponding to a surface impedance well inside the inductive region. The purpose of experimenting in the first place with a guide whose surface impedance was inductive rather than capacitive was to establish the suitability of the measurement technique with this form of guide and to obtain a check *in situ* of the permittivity of the dielectric of which the guide was constructed.

Having carried out that work, guide No. 1 was carefully reduced in cross-section to form guide No. 2, whose diameter finally measured 0·7702 ± 0·0001 in. The tool marks on the surface of the rod were scarcely perceptible, and much less than 0·0001 in in depth, so that they were considered to have negligible influence on the performance of the guide. The extreme care with which this guide was made ensured that its surface impedance came within the comparatively narrow limits of special interest.

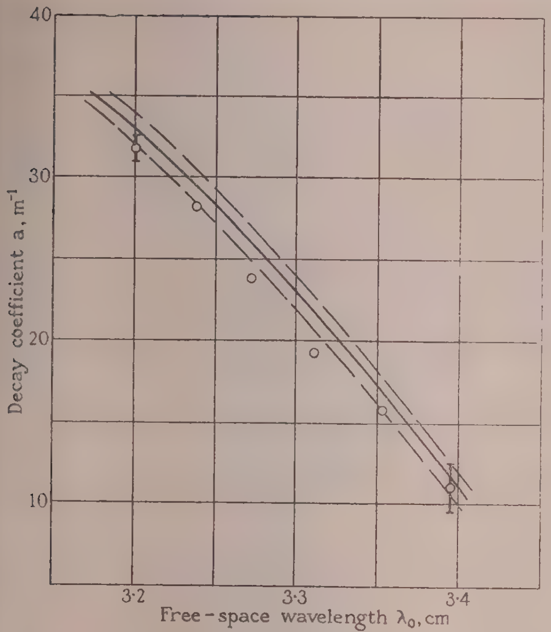


Fig. 6.—Calculated and measured values of the field decay coefficient for guide No. 1, as derived from the resonant frequency and the guide wavelength.

Guide length = 116·91 cm.
Guide radius = 1·0376 cm.

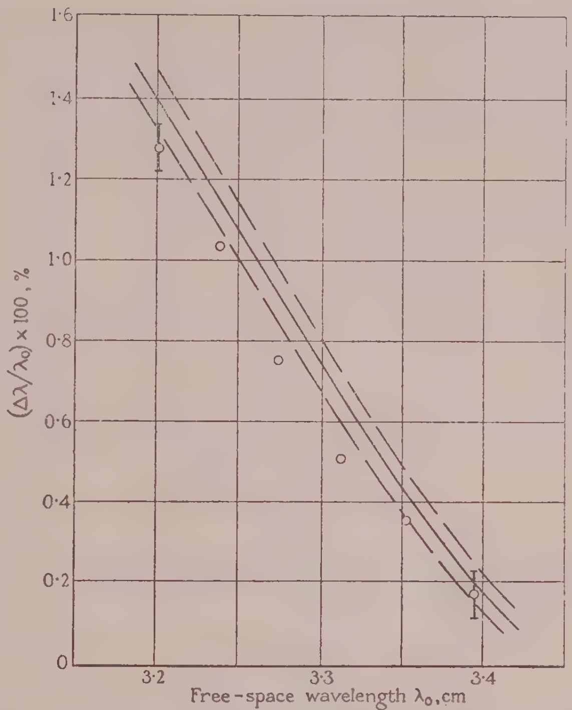


Fig. 7.—Calculated and measured values of the percentage decrease in wavelength along guide No. 1 in relation to the free-space wavelength.

Guide length = 116·91 cm.
Guide radius = 1·0376 cm.

(3.2) Experimental Results and Their Analysis

(3.2.1) Method.

The experimental procedure for determining the properties of a surface waveguide by the resonator technique has been given in detail by the authors[1] in a separate paper. For the present application the approach has been modified in a few respects so as to meet the needs of this particular problem.

The guide to be investigated (guide No. 1) was placed in the surface-wave resonator, as shown in Fig. 4. The length of the resonator was in all cases about 4 ft, and as has already been demonstrated,[15] it could not be made much smaller without appreciably interfering with the purity of the surface wave. The resonator was then carefully aligned, and a thin nylon thread was used to prevent the guide from sagging. The microwave supply frequency was adjusted to establish resonance, as indicated by the output from the crystal detector of the radial-field-measuring instrument, and the resonant frequency was noted. This was repeated at as many resonant points as possible (Fig. 7) with the particular valve generator (CV129) employed. Each measurement of a resonant frequency was accompanied by a careful determination of the resonator length to within 5 parts in 10^4. Finally, from a knowledge of the resonant frequency, $f_0 = c/\lambda_0$, and the length of the resonator, the decay coefficient* a was calculated. With the help of the radial-field-measuring instrument, an experimental field distribution curve (Fig. 5) was plotted when required and compared with the theoretical one appropriate to the guide under consideration.

The above procedure was subsequently repeated for guide No. 2.

A third set of results was obtained using guide No. 2, but with a slightly different length, so that the resonant points fell between those of the original guide (Fig. 13).

* The decay coefficient for guide No. 1 was predominantly real ($|u| \simeq a$) since the guide surface impedance was predominantly inductive.

(3.2.2) Properties of Perspex.

To enable a comparison to be made between the experimental results and those obtained from a theoretical analysis it is necessary to know the electrical properties of the Perspex guide at the frequency employed, and in particular its permittivity must be determined to a high degree of accuracy. With that objective, a number of independent measurements were made on behalf of the authors by C. M. J. Watkins.

A guide method (described briefly in Section 7.1) at $\lambda_0 = 3 \cdot 2$ cm utilized a sample cut from the same Perspex rod as was employed to make guides Nos. 1 and 2. The mean value for ϵ_r, obtained from 12 separate measurements on different sizes of specimen, was $\epsilon_r = 2 \cdot 617$, with a total experimental spread of $\pm 0 \cdot 009$ and a standard deviation of $0 \cdot 005$. The corresponding result for the dielectric loss was $\tan\delta = 0 \cdot 0077 \pm 0 \cdot 0004$.

Other samples cut from a sheet of unplasticized Perspex of material similar to that used for the guides yielded a value of ϵ_r which was, within limits of experimental errors, the same as that obtained for the samples taken from the guide itself.

For the purpose of calculating the theoretical results, the electrical properties of Perspex at $\lambda_0 = 3 \cdot 2$ cm were assumed to be

$$\left.\begin{aligned} \epsilon_r &= 2 \cdot 617 \pm 0 \cdot 007 \\ \tan\delta &= 0 \cdot 0077 \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (27)$$

(3.2.3) Examination of Experimental Results.

(3.2.3.1) *The Identification of the Surface Wave.*

In order to identify a surface wave, a comparison was made between the radial field distribution curve (Figs. 5, 10A, 11A

and 12) obtained experimentally and the corresponding theoretical curve. This comparison is accomplished with the help of an "identifying line," by the procedure previously described[1] in detail for the case in which u is predominantly real. For the present experiments it was necessary to employ values of u which were complex and whose imaginary parts often exceeded the real parts. The procedure therefore had to be modified as follows.

Let $f(x)$ be a function of x that depends on a coefficient ν in the following manner:

$$f(x) = AF(\nu x) \qquad (28)$$

where A is a constant. The coefficient ν (if real) can be found as follows: Suppose $x_1 < x_2$ and let the "identifying constant" be n

where
$$n = \frac{x_2}{x_1} \qquad (29)$$

with "identifying ratio"

$$= p_n(X_1) = \frac{F(\nu x_2)}{F(\nu x_1)} = \frac{F(X_2)}{F(X_1)} \qquad (30)$$

then we can plot an "identifying curve" representing $p_n(X_1)$ against X_1, over any required range, for a given value of n.

If ν is complex we assume the following definition of the identifying ratio:

$$p_n(X_1) = \frac{|F(\nu x_2)|}{|F(\nu x_1)|} = \frac{|F(X_2)|}{|F(X_1)|} \qquad (31)$$

where $ph(X_1) = ph(\nu)$, and eqn. (31) represents an identifying curve of $|F(X)|$ for a given value of $ph(\nu)$.

The coefficient ν as a complex quantity is evaluated as follows: From the graph of $|f(x)|$ we tabulate the ratio $|f(x_2)|/|f(x_1)|$ against x_1 for a suitable value of n. Then it is obvious that, if $ph(X_1)$ has been chosen correctly,

$$\frac{|f(x_2)|}{|f(x_1)|} = p_n(X_1)$$

so that if we read off from the identifying curve corresponding values of $|X_1|$ and plot them against x_1 we obtain a straight line (identifying line) passing through the origin, whose slope is $|\nu|$, whilst $ph(\nu)$ is equal to the phase angle of X_1. On the other hand, if the identifying line does not turn out to be a straight line the phase angle of X_1 has been chosen incorrectly, and another value must be tried.

The functional identification and the evaluation of a complex parameter, ν, as described above may seem a lengthy process, but in practice, once an appropriate set of identifying curves have been obtained, the procedure is quite quickly carried out, as will be seen by the following example.

The readings of the radial-field-measuring instrument are proportional to the magnetic field strength. Thus, if the field inside the resonator represents a pure surface wave, the readings of this instrument are, as a function of radial distance r, represented by

$$f(r) = A|H_1^{(1)}(jur)| \qquad (32)$$

We choose a convenient identifying constant $n = 2$, say,

so that
$$n = 2 = \frac{r_2}{r_1} \qquad (33)$$

and
$$p_2(X_1) = \frac{|H_1^{(1)}(jur_2)|}{|H_1^{(1)}(jur_1)|} = \frac{|H_1^{(1)}(jX_2)|}{|H_1^{(1)}(jX_1)|} \qquad (34)$$

A family of identifying curves is now plotted (Fig. 8), and these are graphs of $p_2(X_1)$ against $|X_1|$ with $ph(X_1) = ph(u)$ as a parameter.

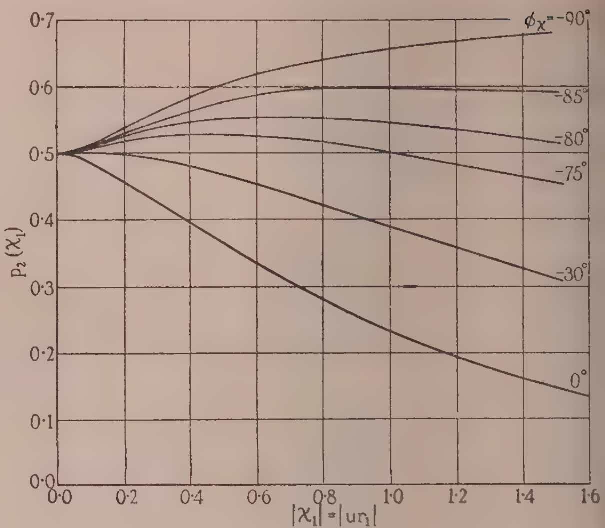


Fig. 8.—Identifying curves for the function $|H_1^{(1)}(jur)|$.

Identifying ratio $= p_n(X)_1 = \dfrac{H_1^{(1)}(jX_2)}{H_1^{(1)}(jX_1)}$

Identifying constant $= n = \dfrac{X_2}{X_1} = \dfrac{r_2}{r_1} = 2$

The decay coefficient, u, is evaluated as follows. Consider the experimental field distribution curve shown in Fig. 5. We tabulate from the curve the ratio $f(r_2)/f(r_1)$ against r_1 for $n = 2$, and since this must be equal to $p_2(X_1)$, we refer to Fig. 8 to find the corresponding value of X_1 for a chosen angle $ph(X_1)$ (say 0°). We then plot $|X_1|$ against r_1 to obtain a straight line (Fig. 9)

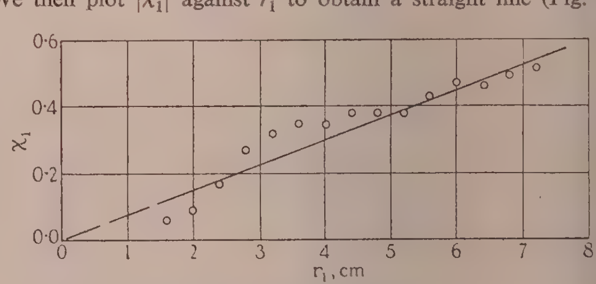


Fig. 9.—Identifying line for the field-distribution curve in Fig. 5.

Measured slope of line $= 7 \cdot 6\,\mathrm{m}^{-1}$

$ph(X_1) = 0°$.

Theoretical value of decay coefficient $= 10\,\mathrm{m}^{-1}$.

that should pass through the origin. Unless the phase angle of X_1 has been chosen incorrectly, the linearity of the identifying line is a measure of the purity of the wave, and its slope gives the value of $|u|$.

Figs. 10A, 11A, and 12 show three different decay curves, and their respective identifying lines are in Figs. 10B and 11B, all referring to the same guide but obtained at different frequencies.

(3.2.3.2) *The Guide Wavelength.*

A direct method of determining the properties of dielectric guides at a given frequency would be to have a number of guides made of the same material but of different diameters and then to measure the guide wavelength for surface waves supported by each of the guides in turn, at the same frequency. Since the constants of Perspex (μ, ϵ_1 and σ_1) do not vary appreciably with frequency, it is more convenient and equally useful to investigate one guide at a number of different frequencies, as discussed above.

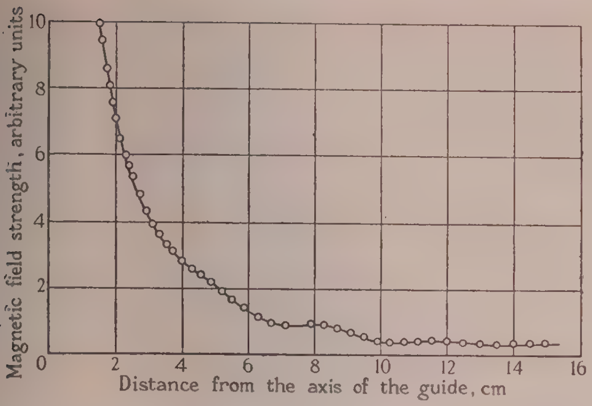


Fig. 10A.—Magnetic-field-distribution curve for guide No. 2, and giving point D in Fig. 13.

Guide length = 116·90cm.
Guide radius = 0·97815cm.
$\lambda_0 = 3\cdot1320$cm.

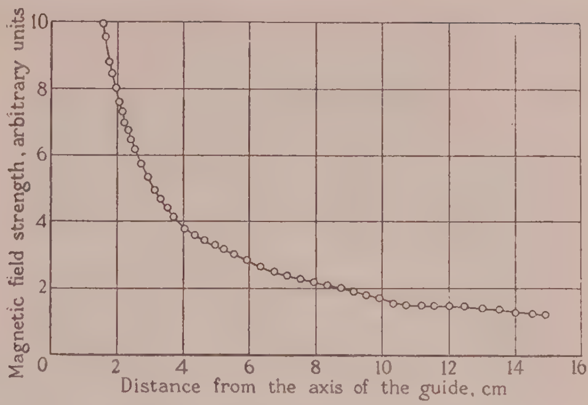


Fig. 11A.—Magnetic-field-distribution curve for guide No. 2, and giving point B in Fig. 13.

Guide length = 116·90cm.
Guide radius = 0·97815cm.
$\lambda_0 = 3\cdot2452$cm.

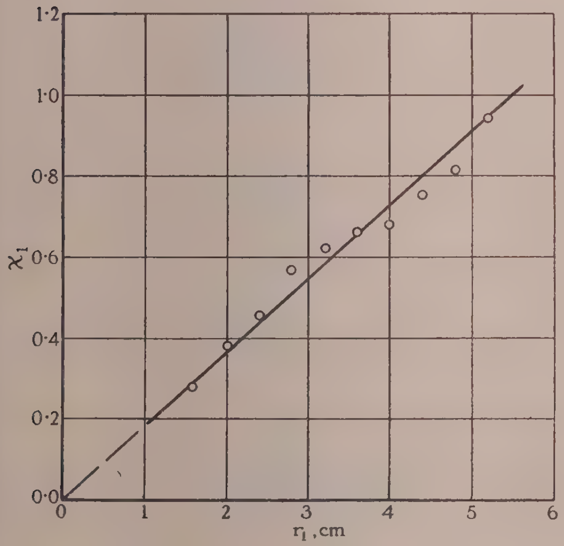


Fig. 10B.—Identifying line for the field-distribution curve in Fig. 10A.

Measured slope of line = 19·4m^{-1}.
$ph(\chi_1) = 0^\circ$.
Theoretical value of decay coefficient = 20·6m^{-1}.

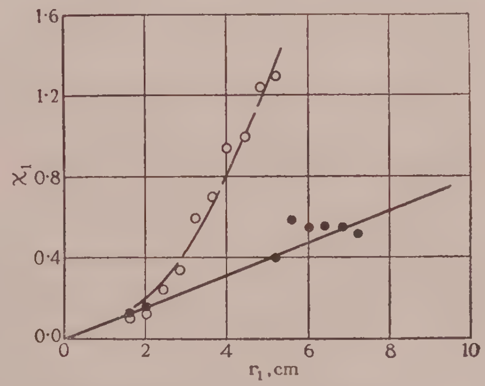


Fig. 11B.—Identifying line for the field-distribution curve in Fig. 11A.

○ $ph(\chi) = 80^\circ$.
● $ph(\chi) = 75^\circ$.

Measured slope of line = 8·0m^{-1}.
Theoretical value of decay coefficient = $u = 6\cdot8\angle 55^\circ$m^{-1}.

The quantities measured are the resonant frequency $f_0 = c/\lambda_0$ and the guide wavelength λ_g. These two quantities enable us to calculate the decay coefficient, a, provided that the guide surface is predominantly reactive, e.g. guide No. 1 (Fig. 6).

In general it can be shown that

$$\frac{1}{2}\tau \simeq \frac{\Delta\lambda}{\lambda_0} = \frac{\lambda_0 - \lambda_g}{\lambda_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

The percentage decrease in wavelength of the surface wave, i.e. $(\Delta\lambda/\lambda_0)100$, is an important quantity and has been plotted from the experimental information for guides Nos. 1 and 2 (Figs. 7 and 13), together with the theoretical curves. These theoretical curves were calculated using the formulae in Section 2 and the constants given in eqn. (27).

The experimental errors, as indicated in Figs. 6, 7 and 13,

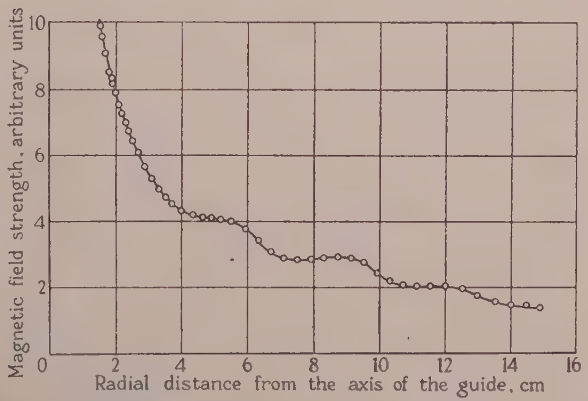


Fig. 12.—Magnetic-field-distribution curve for guide No. 2, and giving point C in Fig. 13.

Guide length = 116·90cm.
Guide radius = 0·97815cm.
$\lambda_0 = 3\cdot2868$cm.

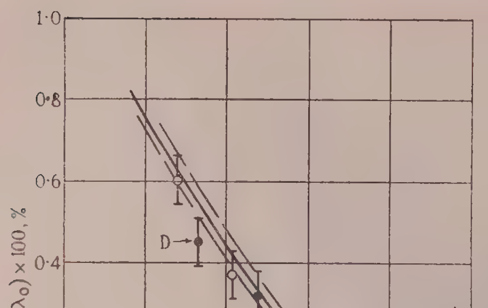


Fig. 13.—Calculated and measured values of the percentage decrease in wavelength along guide No. 2 in relation to the free-space wavelength.

Guide radius = 0·9781(5) cm.

● Experimental points for guide length of 116·90 cm.
○ Experimental points for guide length of 116·28 cm.

have been calculated on the assumption that the error in λ_0 is ± 1 part in 10^4 (wavemeter errors) and that the guide wavelength λ_g has been determined with an error not exceeding 5 parts in 10^4.

The field-distribution curves shown in Figs. 5, 10A, 11A and 12 correspond respectively to points A, D, B, and C in Fig. 13.

(4) CONCLUSIONS

Previous experiments by the authors[1] on the use of a single isolated conductor as a waveguide left no room for doubt that the axial cylindrical surface wave ($E_{0\times}$ mode) is a physical reality when the impedance of the surface of the guide is inductive. We should expect that if the radius of the guide were increased indefinitely the field distribution would ultimately take the form of a Zenneck wave, and at the same time the phase velocity along the surface of the guide would change from a value rather less than the free-space velocity to a value which is slightly greater. Although the excitation of a Zenneck wave propagated over a highly inductive flat surface has been demonstrated experimentally,* there is so far no confirmation that a similar surface lightly loaded is capable of performing the same kind of service, and in fact there has been wide controversy on that subject (see Section 7.2). Very considerable analytical difficulties have been met in attempting to provide an answer to the problem, and the direct experimental approach has proved almost equally inconclusive.

Whilst the authors do not claim to have resolved the difficulty completely, they can justifiably conclude from the experiments described here that there is no doubt at all about the physical reality of an axial cylindrical surface wave having a phase velocity along the guide greater than the corresponding free-space value. It has been established that, not only is a capacitive surface capable of supporting a surface wave of this kind, but, within the limits of experimental error, such waves were found to vanish completely when the threshold of the conditions necessary for maintaining their support in the capacitive region was approached exactly as predicted by theory.

As already explained, the use of a dielectric-rod waveguide was helpful in that it lent itself to following the effect on the surrounding field distribution of the transition from an inductive to a capacitive surface. With the larger-diameter guide, which was inductive, the results (Figs. 6 and 7) conform to the usual pattern expected from calculation, except that they seem to indicate that the bulk permittivity of the Perspex used is 2·610 (the lower limit of the value obtained from measurements on small specimens) rather than 2·617. The scatter of the experimental resonant points is perhaps larger than would normally be expected, but the surface finish of the guide at this stage was not as good as it might have been.

The results of the crucial test are exhibited in Fig. 13, and this shows beyond any doubt that a surface wave whose phase velocity exceeds that of light can be launched. A study of Fig. 13 together with Figs. 5, 10A, 11A and 12 reveals that the surface waves considered in these experiments were not particularly pure; this impurity was not a consequence of the phase velocity exceeding that of light but arose rather from the high attenuation of the guide. Furthermore, much more pronounced impurity became evident just prior to the guide failing to support a surface wave (point C in Fig. 13 and the corresponding curve, Fig. 12).

It has been pointed out that Perspex is not the ideal material as regards its electrical properties for use as a guide, since it gives rise to excessive attenuation. Perspex has a permittivity of the right value and has good machining properties, but its loss tangent ($\tan\delta$) is much too high. A desirable figure for the loss tangent is of the order of 0·004, and whilst a dielectric guide whose loss tangent is less than, say, 0·002 would naturally provide for a purer surface wave, the precision with which such a guide would have to be machined, to obtain a capacitive surface impedance of the right order, would be impossible to attain in practice.

In general, the impurity content of the wave prevented close agreement between the experimental value of the radial propagation coefficient, u, as deduced from the decay curve and the corresponding theoretical value. The purity of the wave improves as the decay coefficient decreases, but when the decay coefficient is very small (Fig. 12), a new source of impurity becomes pronounced owing to diffraction of the wave at the edges of the surface-wave resonator.

It might be argued that the analysis of the field-distribution curves, by reason of their impurity content, should be rejected, but even so, there is still no escape from the evidence supplied by Fig. 13, which is in quantitative agreement with the theory. It was found possible to get resonant points for the system to the left of the line A–A but not to the right of that line. This corresponds exactly to the statement that the operative range of $ph(u)$ is 0 to $-90°$.

These experiments show, therefore, that an axial cylindrical surface wave can be launched over an appropriate guide for all values of the radial propagation coefficient, which, in relation to the surface wave, satisfies Maxwell's equations. This conclusion is drawn in spite of the fact, which is clearly manifest, that a perfectly *pure* surface wave cannot be launched because the field must satisfy Sommerfeld's radiation condition at an infinite distance from the origin. Furthermore, there is little doubt that a ring of magnetic-current filaments concentric with the guide would launch a surface wave over any cylindrical surface for which $ph(u)$ lay anywhere in the range 0 to $-90°$.

* Experiments now in progress at University College, London, have established that radial cylindrical surface waves can be supported by flat metal surfaces thickly coated with Distrene.

(5) ACKNOWLEDGMENTS

The work described was carried out as part of the programme of the Radio Research Board. This paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research. The authors wish to thank their colleagues Dr. A. L. Cullen and Mr. H. G. Effemey for many helpful suggestions in connection with this work and to acknowledge the valuable contributions of Mr. A. H. E. Saxton in making the waveguide to the fine tolerances required, and of Mr. C. M. J. Watkins in measuring its electrical characteristics with great skill and care. They are also indebted to Messrs. H. M. Paisley, E. Rushton and J. V. L. Parry of the National Physical Laboratory for their help in examining the physical properties of the Perspex used for the guide.

(6) REFERENCES

(1) Barlow, H. E. M., and Karbowiak, A. E.: "An Investigation of the Characteristics of Cylindrical Surface Waves," *Proceedings I.E.E.*, Paper No. 1462 R, November, 1953 (**100**, Part III, p. 321).

(2) Barlow, H. E. M., and Cullen, A. L.: "Surface Waves," *ibid.*, Paper No. 1482 R, November, 1953 (**100**, Part III, p. 329).

(3) Bouwkamp, C. J.: "On Sommerfeld's Surface Wave," *Physical Review*, 1950, **80**, p. 294.

(4) Zenneck, J.: *Annalen der Physik*, 1907, **23**, p. 846.

(5) Rice, S. O.: *Bell System Technical Journal*, 1937, **16**, p. 101.

(6) Wise, W. H.: *ibid.*, p. 35.

(7) Whitmer, R. N.: *Proceedings of the Institute of Radio Engineers*, 1948, **36**, p. 1105.

(8) Whitmer, R. N.: *Journal of Applied Physics*, 1952, **23**, p. 949.

(9) Tai, C. T.: *ibid.*, 1952, **22**, p. 405.

(10) Cullen, A. L.: "The Excitation of Plane Surface Waves," *Proceedings I.E.E.*, Monograph No. 93 R, February, 1954 (**101**, Part IV, p. 225).

(11) Ott, H.: *Zeitschrift für Naturforschung*, 1953, **8a**, p. 100.

(12) Burrows, G. R.: *Proceedings of the Institute of Radio Engineers*, 1937, **26**, p. 219.

(13) Barlow, H. E. M., and Cullen, A. L.: "Microwave Measurements" (Constable, 1950), p. 280.

(14) Essen, L.: "The Measurement of Balanced and Unbalanced Impedances at Frequencies near 500 Mc/s and its Application to the Determination of the Propagation Constants of Cables," *Journal I.E.E.*, 1944, **91**, Part III, p. 84.

(15) Barlow, H. E. M., and Karbowiak, A. E.: "An Experimental Investigation of the Properties of Corrugated Cylindrical Surface Waveguides," *Proceedings I.E.E.*, Paper No. 1625 R, May, 1954 (**101**, Part III, p. 182).

(7) APPENDIX

(7.1) Measurement of Relative Permittivity and Dielectric Loss of Perspex

The principle of the method employed to measure the relative permittivity of a sample of Perspex, cut from the rod used as a guide in the surface-wave experiments, is well known,[13] but some modification of the usual arrangements was found to be necessary to meet the requirement of high accuracy in this application. A short length, l, of the Perspex was cut to fit inside the guide of a standing-wave indicator operated in the 3 cm waveband and terminated by a short-circuiting plunger introduced into the end of the guide. This eliminated any couplings and errors that might arise from them. With the probe of the standing-wave indicator set in a convenient fixed position, the guide wavelength with the sample of Perspex removed was first measured by moving the plunger towards the source and observing the distance between two minima recorded on the probe.

The Perspex was then inserted in the standing-wave indicator, and by pushing the sample towards the source with the short-circuiting plunger, a new position of minimum field at the fixed probe was obtained. Provided that the length of the sample of Perspex is small and approximately equal to an integral multiple of half the guide wavelength, the input impedance of the air-filled guide is thus accurately re-established with the Perspex in position so that ϵ_r can be readily calculated. Since the probe was kept in the same position throughout the tests no errors were introduced by variations of probe penetration, whilst the position of the plunger could be read to an accuracy at least of $\pm 0 \cdot 0005$ cm. In this way two separate samples of Perspex were examined, and three independent measurements were made on each, with the incident field first on one face of the Perspex and then on the other.

In calculating ϵ_r a transcendental equation of the form $f(p) = (\tan p)/p$ has to be solved. For finite losses in the dielectric p is complex with a small imaginary part and is represented by $p = \beta l + j\alpha l$. Thus, using Taylor's expansion theorem and introducing $\partial/\partial p[f(p)]$, we find that the error in $f(p)$ arising from the imaginary part of p is a minimum when the length of the sample, l, is an integral multiple of $\lambda_g/2$.

By using the value of ϵ_r obtained as described, and carrying out another series of measurements of open- and short-circuit impedances employing a method similar to that described by Essen[14] in its application to cables, a corresponding value of $\tan \delta$ was derived to a rather smaller degree of accuracy.

(7.2) Conditions required for Launching and Support of a Surface Wave

There is no doubt that a great deal of confusion has arisen regarding the conditions required for the physical existence of the Zenneck surface wave.[3,4] Rice[5] and Wise[6] claim to have established mathematically that such a wave does not form part of the field generated by a vertical dipole when energized and placed above the surface of a plane earth. On the other hand, a number of papers[7–10] have been published showing analytically that a plane surface wave is in fact launched by a vertical dipole located over a loss-free reactive surface. Moreover, it has been demonstrated experimentally that an axial cylindrical surface wave can be supported by a variety of different guides, but in all cases, prior to the present work, the phase velocity of such waves was found to be less than the velocity of light. This feature of the axial cylindrical surface wave which apparently distinguished it from the Zenneck wave seemed to indicate a radical change in its behaviour, and it was this implication coupled with the work of Rice and of Wise that raised the important issue as to whether there was a gradual or a sudden change of guiding properties in the transition from one form of surface wave to the other. It was therefore particularly relevant to inquire whether an axial cylindrical surface wave could be made to travel faster than light and whether the phase velocity or the position of the pole that occurs in the complex plane of the propagation coefficient has any particular significance in the matter. It is of interest to note that in a recent paper,[11] Ott claims to have disproved the conclusion reached by Wise.

In formulating an analytical solution to the launching problem of the surface wave (from a line source in the case of a plane surface wave, and from a ring source in the case of the cylindrical surface wave), we arrive at an integral solution to Maxwell's equations, and the difficult part of the analysis is the exact interpretation of the integral expression. This integration is carried out in the complex γ-plane (Fig. 14), and the important consideration is the presence of two branch points, B_1 and B_2,

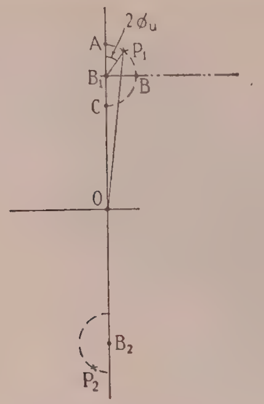


Fig. 14.—Complex γ-plane.

and two poles, P_1 and P_2. The branch points are located on the imaginary axis as indicated in Fig. 14, and their position is independent of the guide surface impedance. The position of the pole, on the other hand, is determined by the value of Z_s; and in particular if $|Z_s| < Z_0$, $|u| < |K_0|$ and the position of the poles is given by

$$\pm j\sqrt{(u^2 - K_0^2)} \simeq \pm j\frac{2\pi}{\lambda_0}\left[1 + \left(\frac{\lambda_0}{2\pi}\right)^2 \frac{|u|^2}{2} \angle 2\phi_u\right] \quad . \quad (36)$$

Thus for constant $|u|$, the pole in the upper half-plane is always located on the arc of a semicircle centred at the branch point and of radius $(\lambda_0/2\pi)^2|u|^2/2$. As ϕ_u passes through all values from 0 to $-90°$ (while $|u|$ is constant) the pole describes the arc ABC. When $\phi_u = -45°$, the pole is at B, so that in this case $\mathscr{I}(\gamma) = \beta = -jK_0$ and the surface wave (if launched) travels at a phase velocity substantially equal to that of light.

The problem involves integration along the whole of the imaginary axis (from $+\infty$ to $-\infty$), and the expression obtained is subsequently transformed into a contour integral by closing the path of integration with a large semicircle (radius $\rightarrow \infty$) as shown in Fig. 15 and applying Jordan's lemma. The result of

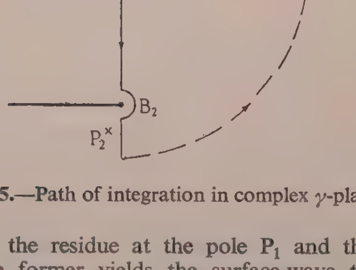


Fig. 15.—Path of integration in complex γ-plane.

integration is the residue at the pole P_1 and the branch-cut integral. The former yields the surface-wave term and the latter furnishes the radiation field plus possibly a surface-wave term. Thus, in some circumstances, unless the branch-cut integration can be carried out accurately, we have no conclusive information about the existence of the surface wave. The position of the pole in the plane of γ plays an important part in the solution of the problem. In particular, it is easy to show by application of the method of steepest descent that, if the pole P_1 is on the imaginary axis $[ph(u) = 0]$ the asymptotic solution to the problem is a surface wave, but for any other position of the pole the field at large distances from the launching point is substantially the radiation field.

With the plane surface wave the coefficient u is directly proportional to Z_s, and we have

$$u = -K_0\left(\frac{Z_s}{Z_0}\right) \quad . \quad . \quad . \quad . \quad . \quad (37)$$

so that $ph(u) = -90° + ph(Z_s)$. Thus if we keep $|Z_s|$ constant and vary $ph(Z_s)$ through all the values from 90° to 0°, the pole will describe the arc ABC. For example, when $ph(Z_s) = 45°$ (conducting earth) the pole is at B and the wave travels with a phase velocity substantially equal to that of light. However, for predominantly reactive surfaces $[ph(Z_s) > 45°]$ the pole is on the arc AB and the corresponding phase velocity is less than the velocity of light. On the other hand, when the pole is on the arc BC (predominantly resistive surface, e.g. dielectric earth) the surface wave travels with a phase velocity in excess of that of light.

With the cylindrical surface wave the relationship between Z_s and u is given by eqn. (3) instead of eqn. (37), so that for any given value of Z_s the angle ϕ_u is less than it is with the plane surface wave. Thus the corresponding pole for a cylindrical surface wave is located on a semicircle ABC of a larger radius than is appropriate to a plane surface wave, and the angular position of the pole $2\phi_u$ is smaller with the cylindrical form of the wave. For this reason the cylindrical surface wave when supported by a plain conducting wire travels with a phase velocity less than that of light, and to obtain conditions resembling those of the Zenneck wave (when supported by the earth at a frequency of about a few megacycles per second) a capacitive surface must be provided.

It is concluded, therefore, that a surface wave can exist for *all* positions of the pole to the right-hand side of the imaginary axis in the complex plane of γ. The experimental evidence strongly suggests that it must be possible to launch a Zenneck wave over a plane earth, including the case of the dielectric earth (almost purely resistive surface impedance). At the same time, there is no doubt that the "efficiency" of launching a surface wave is a function of the surface reactance, and the smaller the reactive part of Z_s (other factors being equal) the lower is the efficiency of launching. Moreover, the maximum surface-wave purity that it is possible to achieve is limited by the magnitude of the resistive part of Z_s, and other factors being equal, the larger R_s the higher the guide attenuation, and consequently the smaller will be the maximum attainable purity.

Although, as the authors firmly believe, a Zenneck wave is present in the field produced by an oscillating dipole over a plane earth, it cannot be distinguished by simple measurements on account of the large impurity content of the field. Indeed, the proportion of the Zenneck wave in such circumstances must be very small, and only highly discriminating measurements could be expected to detect its presence.* The series of experiments described in the paper were designed to circumvent these difficulties and provide the necessary evidence in a different way.

* For this reason the measurements due to Burrows[12] should perhaps be regarded as inconclusive in giving evidence for or against the presence of the Zenneck wave, which must in the circumstances have been of very small magnitude.

621.317.7.087.5 : 778.3

Paper No. 1696 M
Aug. 1954

EQUIPMENT OF INSTRUMENTAL ACCURACY FOR RECORDING AND REPRODUCTION OF ELECTRICAL SIGNALS, USING CINEMATOGRAPHIC FILM

By H. McG. ROSS, M.A., Associate Member.

(The paper was first received 9th March, 1954, and in revised form, 6th May, 1954. It was published in August, 1954, and was read before a joint meeting of the Measurements *and* Radio Sections *4th January, and the* Northern Ireland Centre *19th April, 1955.)*

SUMMARY

Equipment has been made for the recording of electrical signals and their reproduction or play-back, providing an accuracy comparable with that of typical electrical instruments. The errors are about ±1% over most of the range, but increase at the maximum amplitude and frequency; the frequency range is 0–1 kc/s. The equipment is primarily intended for the recording of experiments of short duration or a non-repetitive nature and their subsequent reproduction for study or analysis; a particular feature is the facility for reproduction on a changed time scale, which may be faster or slower than real time.

It is thought that the apparatus offers a significant addition to experimental technique, particularly when used in conjunction with an electronic analogue computer.

Cinematographic methods are employed. A simple film-traversing mechanism is used, with a variable-speed drive. Appropriate optical units are attached to it for recording or reproduction. One type of recording unit gives variable-density tracks on the film; it uses variable-intensity discharge lamps the outputs of which are monitored by photocells, to provide negative feedback to the amplifier. Precision photographic processing methods have been evolved for this application. An alternative optical unit gives variable-width tracks, using modified Duddell oscillographs and being based on standard cinematographic practices, including processing. This unit was made to determine the relative suitability in use of the two techniques, and practical experience shows that it gives significantly greater accuracy, constancy of calibration and reliability in operation. Subsidiary equipment provides for recording on film, mechanical signals, data transcribed from other recording apparatus or numerical data applied manually; this is of value for certain types of analysis using the analogue computer. The importance of making the equipment suitable for use by non-specialist operators has been recognized during the design period.

On reproduction from the records, the light transmitted through each track is measured with a photocell, the outputs from the equipment being presented as voltages; the film provides two tracks carrying independent information of high accuracy, and a third track for timing marks or a reference frequency.

A thorough discussion is given of the considerations underlying the design of the equipment, and results of experiments are presented to show the performance attainable and the types of error which occur.

In Appendix 11.2 a summary is given of the major features of the cinematographic, magnetic tape and disc methods of recording, when used for this application. It is concluded that a comprehensive installation might include equipment of more than one type.

(1) INTRODUCTION

(1.1) General

Equipment has been developed for the recording of electrical signals and their subsequent reproduction or play-back; the apparatus is intended to be an adjunct to electrical measuring equipment, and therefore has to provide an accuracy comparable with that usually associated with typical electrical instruments. Quantitatively, this requirement has been interpreted by attempting to reduce the errors to 1%, while it has been considered that any error as great as 20% would render the equipment virtually worthless. In comparison with the many well-established techniques of sound reproduction this requirement represents a reduction of errors by about one order of magnitude, and the difference is symbolized by the expression of the errors as a few per cent rather than as a few decibels.

The primary application of the equipment is for the recording of the results, expressed as electric signals, of experiments of short duration or of a non-repetitive nature, and for their subsequent reproduction for study or analysis, particularly in conjunction with an electronic analogue computer. Subsidiary applications which have been envisaged include the correlation of the results of some experiments with others taken at another place or time; the reproduction of results which require complex analysis (see Appendix 11.1); or the reproduction of results on a different time-scale, which may be either faster or slower than real time. In certain applications the equipment may be of value as a means of reducing the complexity of the apparatus in use at the time of the experiment, with a consequent net gain in reliability and an increase in the likelihood of obtaining useful results from the experiments.

The equipment is, of course, supplementary to existing methods of recording, usually involving a cathode-ray tube and moving-film camera, which provide no means of reproduction other than visual inspection or point-by-point measurement. However, it is probable that the recording accuracy called for in the equipment described is greater than that normally provided.

(1.2) Review of Earlier Work

The development of this equipment followed an investigation and manufacture of prototypes at the Admiralty Signal and Radar Establishment carried out by H. W. Pout, assisted first by E. N. LeFevre and later by K. W. Thwaites. During the war it became increasingly apparent that weapon-system trials demanded the collection and reduction of large quantities of data which, employing the methods of recording and analysis then available, would have required about 20 man-years for each year of trials. Furthermore, at the end of the war this situation seemed likely to deteriorate as the weapon systems became more complex, and it was concluded that improved techniques would be essential for the planning of large-scale experiments and for the recording and analysis of their results. Early in 1947, consideration was given to the problems of recording such data and no less than thirteen schemes were reviewed, employing various electromagnetic, electrodynamic, electrostatic and optical methods. The most promising technique appeared to be a cinematographic process, based on variable-density records exposed with a new type of controllable discharge lamp. Work had begun on a recorder by the end of 1947 and the first negatives were produced by May, 1948. An intensive programme of development of the lamps followed, in co-operation with the manufacturers,[1] an important feature of which was the design of equipment and the testing of the light output and performance of each variant of the lamp as it was made. The design of the lamps was finalized by mid-1949.

Mr. Ross was formerly with Elliott Brothers (London) Ltd. and is now with Ferranti, Ltd., London.

In addition, a programme was executed in co-operation with a leading cinematographic processing laboratory to establish the conditions which give the highest possible precision of processing of these variable-density records. This was an important factor in making the method usable.

The work with the prototype apparatus was taken to the stage of showing that the system adopted could meet the requirements of this application, and the further development of the equipment, which is the basis of this paper, was begun late in 1950. Particular attention was then given to making the equipment suitable for a wide variety of applications.

(1.3) Choice of Technique for Recording and Reproducing

Three techniques have been surveyed for this type of recording and reproduction of signals—cinematographic, magnetic tape and disc. It is appropriate to consider their particular features, as set out in Appendix 11.2, rather than make any direct comparison between them, and it is probable that in any comprehensive installation it would be desirable to have a variety of recording equipments. In fact, the choice of technique for any particular application is often dependent on one or two overriding considerations, which may be unexpected from the technical point of view. Thus, in sound reproduction, discs are used for radio programmes owing to the immediate and certain access to any item in the record, and to their great storage capacity. In gramophone recording, magnetic tape is used for the master record since it may be readily transported from the place of recording to the factory, and it permits editing and rejection of errors in performance which may occur during the 20 or 30 min recorded on one matrix. For commercial gramophone records, the pressed disc is unrivalled for ease of manufacture in quantity. In the cinema industry, cinematographic film is preferred for ease of editing and synchronization of the sound with the picture.

In the application for which the present equipment was made, the preliminary assessment led to the choice of cinematographic recording. The primary considerations governing this decision were the accuracies required, both with respect to the amplitude of the signal and also the time-scale of the reproduced information; the ability to change the time-scale over a wide range; and the need for simplicity of the equipment for recording, which might be operated by non-specialist personnel. In addition, it was visualized from the start that the work might result in an investigation to determine the degree of accuracy attainable from cinematographic systems when used in connection with the measurement of electric signals instead of with sound recording.

Supplementary requirements of the equipment were the ability to record and reproduce frequencies in the range 0–1 000 c/s, although some increase of error (−10 or −15%) would be acceptable at the maximum frequency; and the provision of two independent recording channels, with a third for recording timing or identification signals, of inferior accuracy of reproduction. Moreover, the recording equipment should be portable and rugged, and suitable for operation from electric supplies of poor stability.

A further part of the equipment was a device which would record data fed into it manually, or mechanically, by the setting of control knobs or use of a keyboard. This would permit the analysis of data available in discrete units, or the combination of such data with the other experimental results, by reproducing such a film in synchronism with one of the experimental records. For example, if the data recorded manually were a sinusoid this would permit the Fourier analysis of the waveform of the original signal by means of an appropriate computer. This apparatus is described in Section 6.

(1.4) Choice of Photographic Method

It is probable that most of the requirements of the equipment could have been met by a system of photographic recording using the binary digital system. An accuracy of 1% requires only relatively few digits to a "word," the upper limit of frequency could probably have been obtained without undue difficulty, and the resolution of photographic film would be sufficient to avoid excessive consumption of film. However, the method was rejected on the grounds of the complexity of the equipment, mainly electronic, required at the time of recording.

Analogue methods of photographic recording, in which the shape or the light-transmission of the record is proportional to the signal, fall into two groups. First, there are standard cinematographic sound-recording techniques or developments from them; these were chosen, since it was possible to utilize much of the experience and facilities offered by the cinematograph industry. Secondly, there are methods using cathode-ray tubes; when the work was begun there were techniques available which permitted a deflectional accuracy of 1% with a cathode-ray tube, even allowing for the distortions arising from optical refraction in the glass face and the residual instability of the high-voltage power supplies. But it was visualized that compensation would have been required for the variation of exposure with the writing speed of the spot, and complexity seemed unavoidable. On reproducing or reading the films, a cathode-ray tube might also have been used. An exceptional deflectional accuracy would still have been required and in addition means to give correction for the inevitable lateral weave of photographic film as it moved through the traversing mechanism. It might have been necessary to introduce some form of lateral modulation or displacement of the spot, at a carrier frequency, with attendant demodulation circuits. An objection in principle is that such a system does not take advantage of the very great amplification of light transmission which is attainable with the photographic process: the light intensity used in reading is similar to that used in recording (the luminous flux is small, and often calls for electron-multiplier photocells); but in a typical photographic system, of reasonable optical efficiency, the luminous flux transmitted through the developed image may be several orders of magnitude greater than the flux required for the exposure.

While improvements have no doubt occurred in the cathode-ray tubes and the circuits for using them, it would seem unlikely that any such system could offer the simplicity which has been obtained with the more ordinary method.

(2) GENERAL DESCRIPTION OF THE EQUIPMENT

The form which the equipment has taken may be described with reference to Fig. 1. A simple film-traversing mechanism is provided, the film being driven by a variable-speed motor of the servo type, electronically controlled. This mechanism is used for both recording and reproducing, appropriate optical units being attached to it.

A variable-density recording system based on the prototype was first made, using the gas-filled discharge lamps as variable-intensity sources. These offer relatively good frequency response and linearity between applied signal and light intensity, and their performance was improved by introducing overall negative feedback, monitoring the light output with a photocell. With this system, precision photographic processing is required. The electronic units have throughout been made in the form of conveniently-sized boxes to permit of easy transportation and of the use of the equipment in confined spaces. The units are connected together by multiway cables.

Variable-area recording optical units were also made, as an alternative, to compare the usefulness of the two systems under

practical conditions when the equipment would be regarded merely as a tool. This system was based on standard practice, using modified Duddell-type oscillographs.

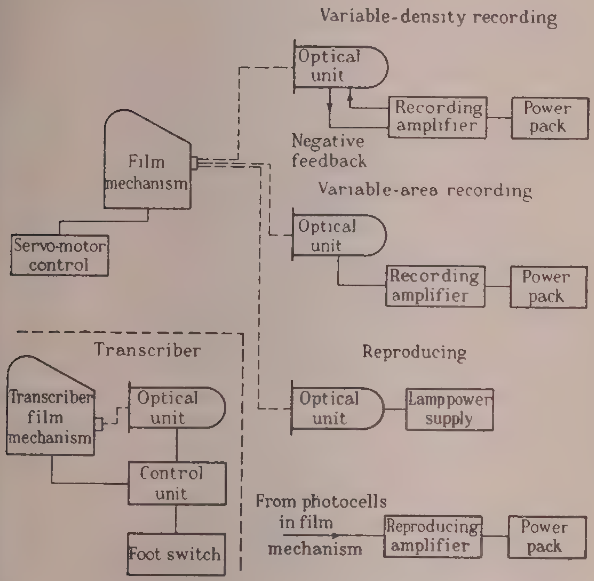


Fig. 1.—Block diagram of components of recording and reproducing equipment.

For reproduction, a simple optical system is attached to the film mechanism, throwing a narrow line of light across the print. Light transmitted through the appropriate parts of the print is detected by photocells, mounted behind the film gate in the mechanism.

A separate film mechanism is used for the recording of data applied manually, providing an adjustable intermittent motion of the film. The data are introduced by setting knobs on an optical unit and are recorded by pressing a foot-switch. Prints from films so exposed are reproduced on the other equipment.

(3) THE FILM TRAVERSING MECHANISM

In order to meet the requirement for uniformity of motion of the film or tape, the design of mechanisms for the movement of film is often based upon the use of a free-running roller or drum, of relatively large inertia. The film is held in contact with the drum and usually serves to drive it. A loop has to be provided on either side of the drum, to isolate any irregularities of motion introduced from the spools, and it is normal to give some form of damping to prevent the development of an oscillatory condition within these loops and the drum. Alternatively, a more direct form of film drive may be used, with a vibration-damping device to eliminate irregularities of motion introduced from the motor or reduction gearing.

In the present application, it was thought that the requirement of the wide range of speeds of movement of the film would make it very difficult to obtain satisfactory operation with such a system; indeed, at certain speeds excessive irregularities might have been introduced. Consequently, the film drive was designed round a standard toothed film sprocket, driven through spur gearing from the variable-speed servo motor.

The equipment provides film speeds with a high degree of constancy over the range 10–1 in/sec, used for recording; all normal variations of load and of voltage or frequency of the electric supply do not give errors greater than 1% at any setting. Continuous adjustment is provided below this speed, down to zero, but the reduced output from the tacho-generator associated with the motor results in a progressive impairment of the constancy of film movement; at 0·02 in/sec the motion still appears uniform, and measurements have not shown errors greater than 3%. In addition, there is a short-period irregularity which is associated primarily with the teeth on the gear-wheel driving the main sprocket-shaft. This can be detected as a slight frequency-modulation of a high-frequency signal or an amplitude modulation of a variable-density recording. In the absence of special test gear, it was measured by exposing film with a narrow transverse line of light and analysing the consequent density fluctuation. The method did not permit great accuracy, but at the maximum film speed it indicated an irregularity of peak amplitude of $\pm 1\%$ and of frequency of about 600 c/s. No irregularity has been observed at a wavelength corresponding to the spacing of the sprocket holes.

Throughout the development of the equipment, the facility for changing the speed of film movement has been invaluable, and it has also been a great simplification to be able to make certain measurements on the film while stationary in the mechanism. Furthermore, at the time of the equipment going into service, the great majority of applications envisaged make use of this facility, the speed on reproduction being either slower or faster than on recording. One particular example of the latter case is the recording of the performance of a mechanical system and its play-back for analysis in an electronic computer, thus to some extent bridging the disparity of the time scales of mechanical and electrical systems.

The mechanism has been designed round commercially available film cassettes and sprocket-shaft assemblies, and has been based on standard cinematographic practices wherever possible, particularly with respect to the angle of wrap of the film on the sprockets, leading the film on to and off the sprockets, and the provision of loops to isolate the film in the gate from disturbances. The cassettes, holding 400 ft of 35 mm film, may be loaded into the mechanism in daylight, and when the upper feed-cassette has been emptied it may be placed in the lower position to take up the next length of film. The take-up drive, which is sometimes a rather hazardous feature of cinematographic equipment, embodies a slipping clutch in which a metal disc is partially gripped between two Oilite discs, adjustably spring-loaded; the design permits ready observation of whether the take-up is working correctly. The system has proved satisfactory. A footage counter is provided.

A straight film gate is used, with the photocells used for reading the film mounted in a block behind it. The straight gate was necessary for the transcriber unit (used for intermittent recording) and its use in the mechanism for continuous recording and reading permitted the same basic design for the two machines. The alternative, of carrying the film past the exposing position on a large-diameter sprocket, might possibly have given slightly less irregularities of film speed, but would have made it difficult to read the film with the photocells.

Jointed film may be run, but no provision is made for using long loops of film which would have needed extra rollers to bring out the film and a holder for the loop. The mechanism does not work in reverse, but means are provided to creep the film backwards, to position a particular point on the record in the gate, although the take-up must then be turned by hand.

Certain computations, for example auto-correlation, may require two mechanisms running in exact synchronism; a small modification permits them to be coupled together mechanically.

(4) THE VARIABLE-DENSITY RECORDING EQUIPMENT

(4.1) General

The preliminary work (see Section 1.2) had established the feasibility of variable-density recording for the present application, using variable-intensity discharge lamps and precision photographic processing. The equipment promised to be small and light in weight, unaffected by mechanical vibration and with simple optical components. Experience also revealed other features of the lamps. First, their light output and linearity were dependent on their temperature, which necessitated—for this application—enclosure of the lamps, circulation of air and thermostatic control at 140° F ± 1° F. Secondly, the light output for a given current was influenced by the previous history of the lamp, and the optimum performance was not often obtained after a life of 100 hours. Thirdly, some lamps would develop irregular operation, corresponding to noise, or would become difficult to strike, although it is understood that this has been made less serious in lamps of recent manufacture.

Consequently, it was decided to develop a system whereby the luminous output of the lamp was monitored by a photocell, which introduced negative feedback to the amplifier providing the current to the lamp. Thus to a first approximation the accuracy of the system was transferred from the lamps to the photocells; other advantages followed, including improved striking and the avoidance of thermostatic control. The price to be paid was a considerable complication of the electronic equipment, so that much of the original simplicity was lost.

(4.2) Variable-Density-Recording Optical Unit

The optical unit for variable-density recording, shown in Fig. 2, is made to be attached to the film mechanism. Each lamp is held kinematically in a ceramic tube which ensures that the lamp reaches a very high temperature (after about 15 min operation) to minimize the possibility of noisy irregularity in the discharge.

The end of the discharge within each lamp is projected by a condenser lens into the objective lens. Adjacent to the condenser lens is an inclined mirror which has a clear line along its length. A line of light from each lamp, of uniform luminosity, is therefore projected on to the film, the luminosity depending on the signal applied to the amplifier. The major part of the light passing through each condenser lens is directed upwards on to separate photocells; the angle of collection of light for monitoring is therefore the same as that for exposure on the film, a point which the preliminary work had established to be important.

Each of the two lamps for recording information is horizontal and inclined slightly to the optical axis in plan view. A third lamp is provided, and to permit the tracks to be close together on the film this lamp is mounted vertically and the light path from it is brought in by a mirror between the light paths of the two other lamps. This comparison lamp, the light output of which is not modulated, is used in conjunction with features in the reproducing system to give some cancellation of any effects which influence the whole film or the modulated and unmodulated lamps together; it is a most important feature of the system, and it is extremely doubtful whether without it adequate performance from the variable-density system could have been obtained.

A fourth lamp with a rectangular aperture over the end of its discharge is reflected by a mirror into the objective lens. This gives the additional track on the film for timing marks or a reference frequency, and not being monitored, gives inferior amplitude accuracy. The three important tracks are each 0·15 in wide and are placed within the standard cinematograph picture area; the timing track is of standard sound-track width and position.

For each speed of film movement used in recording, an appropriate neutral-density filter must be slipped into the optical unit to compensate for the different times of exposure of the film. For recording, fixed speeds of 10, 5, 2 or 1 in/sec are used, and the corresponding densities would nominally be 0, 0·3, 0·7 and 1·0. But the apparent density of a filter is affected by the optical system in which it is used, and an allowance must also be made in this case for the failure of the reciprocity law of exposure of the film. Consequently, the correct densities were determined experimentally, and were 0, 0·25, 0·76 and 1·06, as measured by the manufacturers.

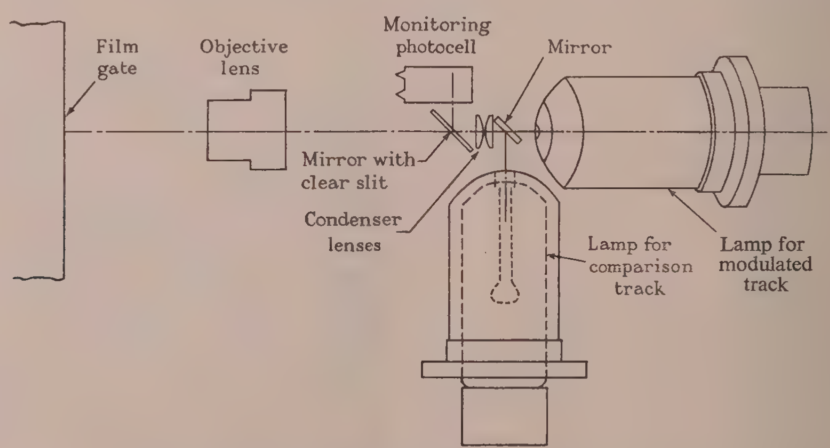


Fig. 2.—Side view of optical unit for variable-density recording.

(4.3) Variable-Density-Recording Electronic Unit

The electronic circuits used with the lamps are given in Appendix 11.3. Each modulated lamp has a cathode-ray tube associated with it to give a visual indication that the range of modulation is not exceeded. At the lower end, over-modulation extinguishes the discharge and the negative-feedback loop is broken; on removing the signal the amplifier applies a very large potential to the lamp which re-strikes it, and this results in a

brilliant instantaneous discharge since it takes about one millisec for the negative-feedback circuit to take control again. At the upper end, over-modulation for any length of time can damage the lamp, but a more insidious trouble, resulting from instantaneous overload, is that the monitoring photocell may suffer a temporary hysteresis effect; at the levels of illumination used, this was found to occur with some, but not all, of the photocells. The oscillographs also serve another most useful function, for they show whether the lamp is suffering from an irregular or noisy discharge; since this is characteristically of 5- to 25-kc/s frequency and of very irregular waveform, the negative feedback does little to minimize it. It may often be stopped by running the lamp for some minutes at a large current, thereby raising its temperature.

(4.4) Performance of Variable-Density Recording Unit

Measurements were made of the performance of the variable-density recording unit alone; this is to be distinguished from the performance of the whole recording and reproducing system (see Section 8.1).

A supplementary photocell was mounted in the optical unit to receive light from the lamp; this was placed just before the condenser lenses and did not interfere with the negative-feedback action. The output from the photocell was measured with a direct-coupled cathode-ray oscillograph, its measuring device having been calibrated with known potentials.

The relationship between the voltage applied to the amplifier and the luminous output of the lamp is shown in Fig. 3(*a*). In this Figure, and others where relevant, the errors or departures from linearity are given, rather than the direct plot; the error is expressed as a percentage of the full range of modulation, although the equipment is normally used with zero at the mid-point of the range (equivalent to a centre-zero meter). The corresponding lamp currents did not show marked non-linearity. The normal range of modulation is 5–120 mA (nominal), which gives a ratio of luminous intensities of 1 : 24, this representing a remarkably wide range for a lamp of any type.

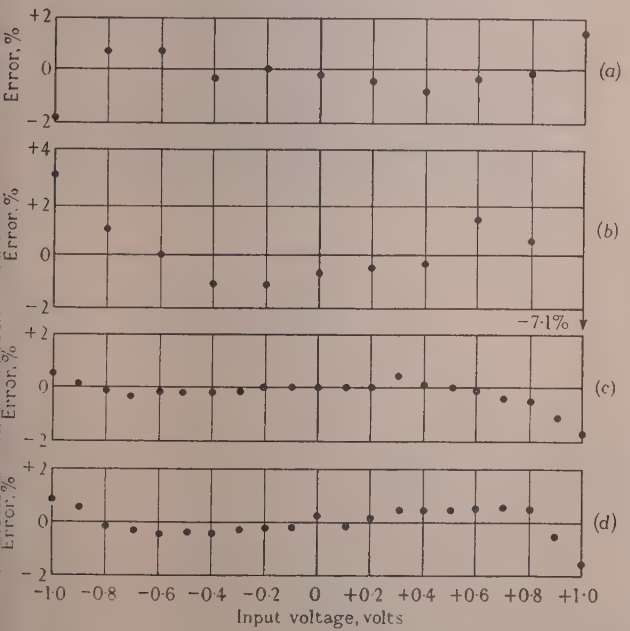


Fig. 3.—Errors in amplitude linearity.

The non-linearity is expressed as a proportion of the full range of recorded signal.

(*a*) Variable-density recording unit alone, showing the difference between input voltage to amplifier and luminous output from lamps.
(*b*) Overall variable-density system, showing the difference between input voltage when recording and output voltage on reproduction.
(*c*) Variable-area recording system alone, showing the difference between input voltage and width of the exposed track on the film.
(*d*) Overall variable-area system, showing the difference between input voltage when recording and output voltage on reproduction.

The frequency response of the recording unit was determined by applying a constant-amplitude sinusoidal voltage to the input, and is shown in Fig. 4(*a*). The time-constants of the circuit and the properties of the lamp resulted in a natural oscillation round the feedback loop at about 3 kc/s; this was eliminated by the relatively large condenser C (Fig. 19), the value of which was chosen so as to leave a rising frequency characteristic to compensate partially for the high-frequency attenuation associated with other parts of the system.

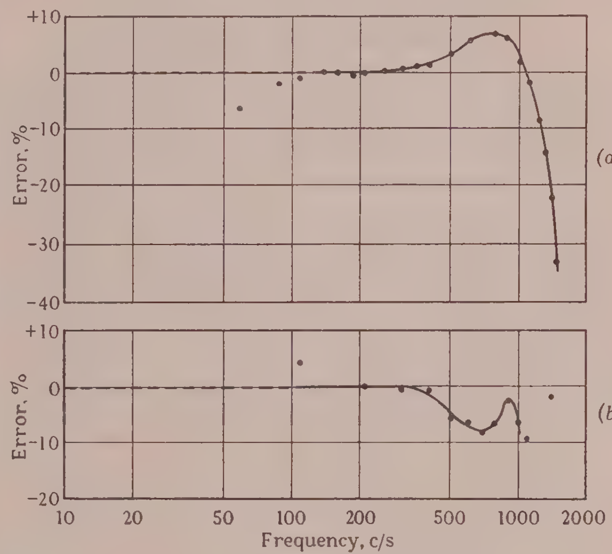


Fig. 4.—Frequency response of variable-density system.

(*a*) Recording lamps and amplifier only.
(*b*) Overall recording and reproducing system. (Film speed: 10 in/sec on recording and 2 in/sec on reproduction.)

(5) VARIABLE-AREA RECORDING EQUIPMENT

(5.1) General

When preliminary consideration was being given to devising equipment for the present application, no method of variable-area recording appeared to show much promise. Every such system relies on an electro-mechanical device, and it was feared that this might be affected by mechanical disturbances; it would be heavy and much larger than the discharge lamps, and would require an optical system which might easily be complicated and of large dimensions. Furthermore, at the time, no adequate measurements were available of the linearity and frequency response of the modulators.

However, such were the advantages of a system which relied on a dimensional effect on the photographic film, rather than on a transmission which requires precision processing, that it appeared to be worth while to develop variable-area recording equipment. It was only later that a secondary advantage was fully realized, there being a very large increase of the quantity of light transmitted through the film during reading, and this improves the performance of the reproducing equipment. The variable-area records are also much easier to interpret visually.

In the event, the variable-area equipment required about as much development effort as the variable-density equipment and is of similar complexity; while both are much more complex than the original equipment which proved the feasibility of the variable-density system. However, at the completion of the development stage the variable-area equipment had been found to give superior results with respect to amplitude accuracy, frequency response and noise level. Subsequent experience under practical conditions showed advantages of simpler operation, greater reliability and superior constancy of calibration.

(5.2) Variable-Area Recording Optical Unit

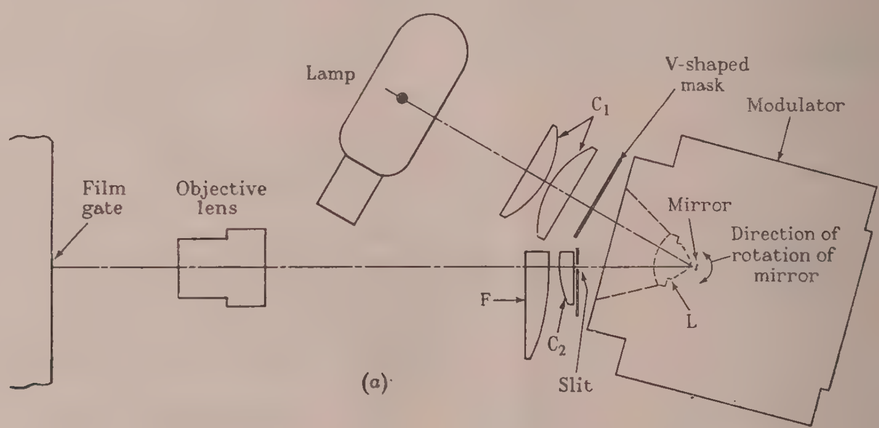
The modulators in the optical unit are Duddell-type oscillographs, developed for one of the modern cinematographic sound-recording systems. The moving element comprises a single loop of duralumin wire, held taut between the poles of a magnet, and carrying at its centre a mirror 0·060in × 0·022in. Below the resonant frequency (normally at about 10 kc/s) this behaves entirely as a simple mechanical system. The whole is filled with silicone fluid. The modulators for this application were modified slightly, and their response to a step function (the best test available at an early stage of the work) showed a time of rise of 1/5 000sec, with no visible overshoot or overdamping. There is no evidence to suggest that they are more microphonic than the valves driving them.

The optical system for the modulators was developed from that designed for sound recording. As shown in Fig. 5(*a*), the lamp is projected by the condenser lenses C_1 on to the mirror of the modulator, and the virtual filament in the mirror is projected by the second condenser C_2 and field lens F into the objective lens. Consequently C_1 and C_2 would appear uniformly and brilliantly lit. A V-shaped mask is held over C_1 and projected by the lens L on to a clear slit formed across an otherwise opaque coating on C_2. This slit is projected by the objective on to the film. When the mirror in the modulator rotates, the image of the mask moves across the slit and the length of the illuminated part of the slit is changed. As the film moves, therefore, a track of variable width is exposed.

The difficult features in the design of the system were to combine on one film the two systems corresponding to the two independent channels of information to be recorded without introducing significant optical distortion, to make the tracks wide enough to use most of the film, and to keep the overall dimensions to a minimum.

The device used is shown in the plan view, Fig. 5(*b*). That part of the optical system from the lamp via the modulator to the condenser C_2 is duplicated, the two optical axes being parallel to, but displaced from, the main optical axis. Thereafter the light passes through the outer parts of the field lens F [shown truncated in Fig. 5(*a*)], and is refracted into the objective lens.

An important feature of the system is that it is possible during manufacture to modify the shape of each V-mask to correct for errors introduced by the optical system or modulators. Thus, following initial tests with a straight-sided mask, an error of 3%



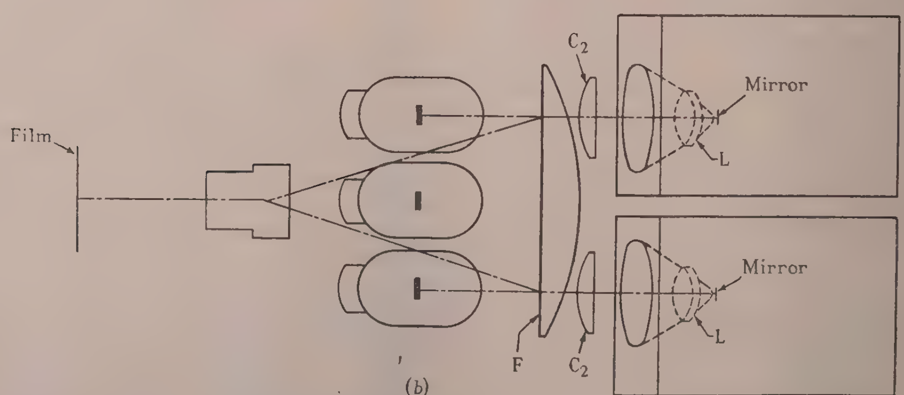


Fig. 5.—Optical unit for variable-area recording.

(*a*) Side view.
(*b*) Part of plan view.

at maximum modulation was reduced, and the relation between the width of exposed track (0·005–0·240 in) and the maximum driving current of ±150 mA was standardized. This makes it possible to interchange optical units and electronic amplifiers (out of four sets) without affecting the overall calibration.

Further optical components (not shown) provide a slit of fixed length, to give on the film the reference track, which is unmodulated. The track for timing or frequency-calibration marks is exposed by one of the variable-intensity discharge lamps, the light from it being reflected into the objective lens. This is little more than an on-off system, with no pretence to amplitude accuracy, since the photographic characteristics are inappropriate for a variable-intensity system.

Three tungsten-filament lamps are used in the optical unit, one for each modulated track and the third for the unmodulated. Although wasteful of power (10 volt, 5 amp each) this greatly simplified the optical design. They are run from the 50-c/s mains supply, but owing to the heavy filament and the use of the comparison track, the residual effects arising from the irregularities of exposure of the film result in errors less than 0·3%.

The appearance of a piece of film on which sinusoidal signals have been recorded is shown in Fig. 6. The footage number on

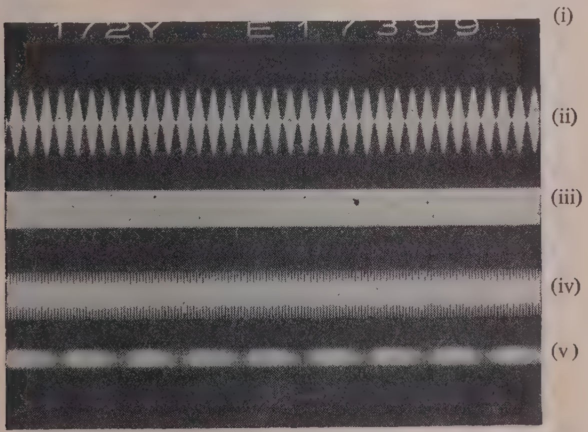


Fig. 6.—Appearance of a variable-area record, of sinusoidal input voltages.

Film speed: 10 in/sec.

Track (i). Footage number, visible on film before and after processing.
Track (ii). 200 c/s, 100% modulation.
Track (iii). Unmodulated or comparison track.
Track (iv). 1 000 c/s, 50% modulation.
Track (v). Timing marks, 50 c/s; variable transmission.

the edge of the film is found to be of the greatest value in this application; it is printed on the film during manufacture, is visible before exposure and also after processing, and it appears on the print. It greatly simplifies noting the records of experiments, and assists in finding the correct place along the film during the subsequent analysis.

Wherever possible, standard cinematographic practices have been followed. This applies particularly to the processing, and it is usual to obtain an over-night service at the processing laboratories, this including development of the negative and production of the print.

In common with other equivalent systems, an attenuation of higher frequencies is caused by the finite dimensions of the exposing slit. The effect is dependent on the size of the slit, in the direction of movement of the film, relative to the wavelength of the exposed trace; it is set out in Fig. 7, and for convenience the corresponding frequencies at two film speeds are added. It occurs on both recording and reproducing, the errors being

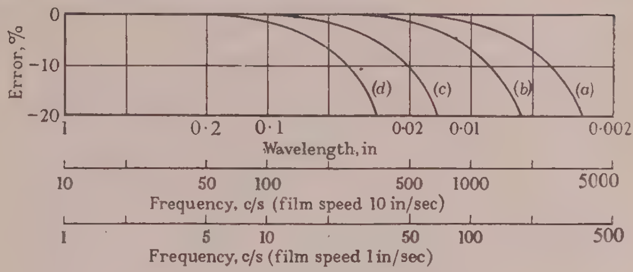


Fig. 7.—Error due to dimension of slit.

Errors in recording and reproduction are additive.
Dimension of slit along film: curve (*a*) 0·001 in; curve (*b*) 0·002 in; curve (*c*) 0·005 in; curve (*d*) 0·010 in.

additive; during recording some even-order harmonic distortion may be introduced, but this and the magnitude of the attenuation are affected by the photographic conditions.

(5.3) Variable-Area-Recording Electronic Unit

Each modulator is of 0·25 ohm resistance. An attempt was made to develop an amplifier using a modulated-carrier system with transformer coupling from the output valves to the modulator, to match the impedances. After transformation, the signal was rectified at a low voltage and relatively large current, and negative feedback was used from the modulator circuit to the input. It was not easy to obtain stability, probably owing to the number of transformers within the feedback loop; the method did not appear promising and was dropped.

In the circuit used, given in Fig. 8, the modulator is connected between the cathodes of a pair of power valves in push-pull. The mismatch is on the grand scale: 160 watts of power is drawn from the mains supply (including the excitation of the lamps) and the maximum power used in each modulator is 5·6 mW. The circuit, however, is unusually simple, each recording channel using only three pairs of valves, and a power pack with metal rectifiers and neon-tube stabilization. The power-supply points shown are connected in common for both the amplifiers to the one stabilized supply; an isolated supply for each recording channel is connected to the points A and B and provides the main power for the output stages. Owing to the feedback connection, the application of in-phase potentials to the input results in the potential of each of these packs with respect to earth being varied at signal frequency; the capacitance to earth of each pack is therefore kept rather lower than usual to prevent impairing the independence of the circuit to in-phase signals.

Negative feedback is derived from precision wire-wound resistors, R_2, in the modulator circuit; about 45 dB is used. The value of R_2 was chosen to give exactly ±150 mA through the modulator for ±1 volt input, thus permitting electronic units to be interchanged on the modulators (in the optical units) without significantly altering the calibration.

The heaters of V_1 and V_2 are stabilized by a barretter. This practice is used for the first stages of all amplifiers of the equipment, and, in conjunction with a minimum input signal of one or two volts (for full-scale deflection), results in adequate zero stability.

(5.4) Performance of Variable-Area Recording Unit

The performance of the amplifier alone was tested with a sub-standard milliameter in series with the modulator, or by replacing the modulator by a resistor and connecting either side of it to a cathode-ray oscillograph which was otherwise isolated.

No errors greater than those expected from the test equipment were observed, the amplitude linearity being tested up to ±180 mA, and the frequency response up to 2 000 c/s.

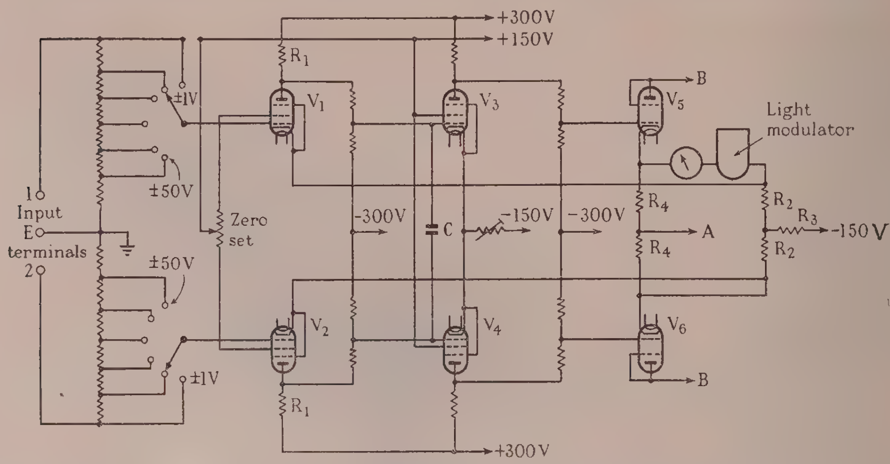


Fig. 8.—Circuit of variable-area recording amplifier.

Negative feedback is derived from resistors R_2. Points A and B are connected to an isolated power supply of 220 volts.
V_1–V_4—EF37A; V_5, V_6—12E1; R_1—150 kilohms; R_2—3·16 ohms precision wire-wound; R_3—75 kilohms; R_4—150 ohms wire-wound; C—0·01 μF.

Concerning the independence to in-phase voltage at the input, an error signal of 0·3% appeared when ±5 volts was applied in phase, with the sensitivity set to the ±1-volt range.

To determine the accuracy of the recorded tracks on film (again to be distinguished from the accuracy of the whole recording and reproducing system) the widths were measured with a travelling microscope. Although, owing to the magnification, the edges of the tracks did not appear truly sharp, measurement by one observer only resulted in the adoption of conventions, and the very small scatter in the observed readings suggests that measurements were relatively correct to within less than 0·2%.

The amplitude linearity is shown in Fig. 3(*c*), where it is seen that the errors only exceed 0·5% at the maximum modulation, corresponding to the widest track.

For measuring the frequency response, the peak amplitude of a sinusoidal voltage applied to the recording amplifier was measured by a backing-off procedure with a direct-coupled cathode-ray oscillograph; the direct voltage was adjusted until the peak of the wave was brought to earth potential. This permitted the use of a sensitive display on the oscillograph, and the actual measurement was made on a sub-standard voltmeter. A reversing switch made possible the measurement of the positive and negative peak amplitudes.

For the experiment of Fig. 9(*a*), the film was recorded at 10 in/sec and frequencies between 1·5 and 2 000 c/s were applied

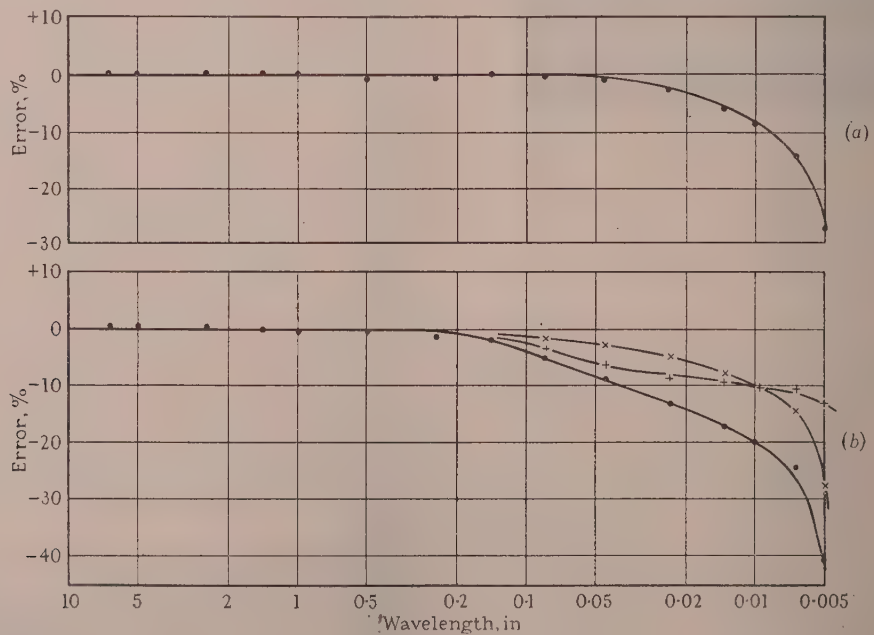


Fig. 9.—Errors due to attenuation at short wavelengths and high frequencies, variable-area system.

(*a*) Recording system alone, the dimensions of the track on the film being measured.
(*b*) Overall system, being the relationship between input voltage when recording and output voltage on reproduction.

—•— Total attenuation.
—+— Attenuation of positive peaks.
—×— Attenuation of negative peaks.

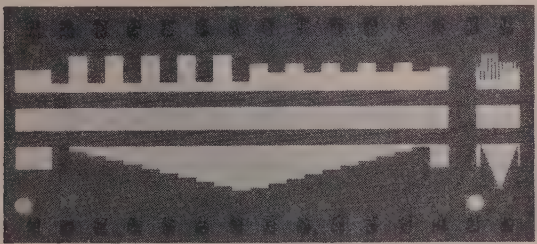
at constant amplitude; the peak-to-peak widths of the modulated tracks on the film were then measured with the travelling microscope. The attenuation is dependent on the wavelength on the film; it is largely attributable to light scatter, but is in part due to the width of the slit in the optical system, which was equivalent to 0·0007 in at the film.

(6) THE TRANSCRIBER OR STEP-RECORDING EQUIPMENT

(6.1) General

Separate equipment has been made for recording on film information applied manually or mechanically. The primary purpose is to provide a means of feeding into the electronic computer—for which all the recording and reproducing equipment has been made—data obtained from other recording instruments (e.g. cine-theodolites), tabulated data, random numbers, the equivalent of noise or jitter, sinusoidal voltages, non-linear functions, the settings of controls of mechanical equipment or records of the operation of mechanical devices. It is a means for transcribing data into a form suitable for the computer.

The equipment comprises a film mechanism, similar to the continuous-film mechanism both outwardly and with respect to the layout of the film path, but differing in that the film is moved intermittently. An optical unit is attached, carrying two pairs of knobs for controlling two tracks of independently recorded information. These may be set manually, in tens and units from 0 to 100, or may be coupled by extension shafts to any external mechanism. The knobs being set, a foot-switch is pressed, a shutter gives an exposure on the film and then the film is moved forward by a distance exactly equal to the length of the exposed patch. The exposed tracks, an example of which is shown in Fig. 10, therefore appear continuous, but the information is



(i)
(ii)
(iii)
(iv)

Fig. 10.—Demonstration film recorded on the transcriber. The same data were recorded at 0·1 in movements and then at 0·01 in movements. Motion is from left to right.

Track (i). Controls set at 50, 50, 20, 80, etc., 60, 40, etc., 50.
Track (ii). Unmodulated or comparison track.
Track (iii). Controls set at 50, 50, 0, 10, 20 . . . 90, 100, 95, 85, 75 . . . 15, 5, 50.
Track (iv). Identification mark track; single exposure at start of each pattern.

recorded on them in definite lengths, and any change in the setting of the controls gives a step. The mechanism may be set to give film movements of 1, $\frac{1}{2}$, $\frac{1}{5}$, $\frac{1}{10}$, $\frac{1}{20}$, $\frac{1}{50}$ and $\frac{1}{100}$ in.

After processing, the print is read in the equipment which is used for the reproduction of the records from the continuous recorders; in general, it is visualized that two sets of reproducing mechanisms will be used simultaneously, one reading the film from the transcriber and the other the film from the original experiment. Therefore the transcriber mechanism is not equipped with photocells behind the film gate, although these could be fitted if a need arose to read a film at definite intervals, after exposure in a continuous mechanism.

It would probably have been better from the operational point of view to control the tracks by pushbuttons, but this would have resulted in a much more complicated mechanism than have the rotating knobs. The disadvantages of the latter are lessened when small differences only occur between successive settings, as is expected to be the case usually.

The equipment is calibrated over the range 0–100, but normally when the film is read the output appears as ± 25 volts, with zero corresponding to the setting 50.

Particularly with short film movements, the exposure of a whole length of film could represent a major task. Consequently, reliability and constancy of calibration are essential, even over periods of several days: the whole apparatus is free of electronic equipment.

(6.2) Transcriber Optical Unit

The optical unit comprises a lamp, operated from a saturated-core stabilizing transformer, projected by a pair of condenser lenses into an objective lens, the condenser therefore appearing of uniform luminosity. Adjacent to this condenser is a system of masks which are projected on to the film by the objective. Two fixed strips define a central track, which corresponds to the unmodulated comparison track of the continuous recorders, and is required for the reproducing equipment. The outermost masks, on either side, are controlled independently by either pair of setting knobs, each unit knob working through a differential screw thread. In addition, one of a set of masks may be dropped in, to define the upper and lower extremes of the exposed areas, the opening of each of these masks corresponding exactly to one of the lengths of film movement.

The shutter, which is over the objective lens, is of the type used in the simplest cameras, having a swinging plate with a hole in it, and a capping blade. It is operated by a solenoid, and a system of relays is used in the control unit to ensure that each exposure is followed with certainty by one film movement, and that the cycle of operations is independent of the manner in which the foot-switch is pressed.

Exposure is made on the fourth track, corresponding to the timing track of the continuous recorders, by switching on a tungsten-filament lamp during one exposing cycle. This illuminates an opal glass, on which identification marks can be written if desired.

It may be noted that an alternative method of using this optical unit is to attach it to the continuous-recording mechanism, and thus obtain a continuous variable-area record of any mechanical motions which have been coupled to the operating shafts.

(6.3) Transcriber Film Mechanism

An error in the length of movement of the film of 0·001 in represents 10% of the shortest movement, and in certain circumstances this can give a significant error in the time scale on reproduction. Such a positional accuracy is the best attainable with special cinematographic equipment and is better than that normal to cinema projectors. This accuracy had to be combined with means for changing the length of movement over a range of 100 : 1, although the error could be greater with the longer movements.

The mechanism employed comprises an accurate ratchet wheel (of 300 teeth, corresponding to a 3 in movement) on the shaft which carries the primary film-driving sprocket, engaged by a pawl on an arm. This arm is actuated by a cam, driven intermittently by a motor of the split-field type often used in servo systems. The fields are constantly energized, but the armature is either connected to a supply, when it runs, or short-circuited by contacts, when it stops suddenly. There is gearing of 1 : 10 ratio between the shaft carrying the cam and the shaft carrying the disc with which the contacts engage. Discs with 10, 5, 2 or 1 notches determine how many times the cam rotates for any one

operation of the foot-switch, and interchangeable cams with a lift of 1 or 10 teeth on the ratchet wheel allow selection of any of the range of movements. Thus the 1 in movement is obtained by ten steps, each of $\frac{1}{10}$ in, after which the motor stops.

The 1 in movement takes the longest time, a complete cycle occupying 2·5 sec.

The system is arranged so that when the motor has stopped the minimum radius of the cam is engaged with the arm, the cam being made circular in this region so that slight irregularities in the stopping position of the motor do not matter.

Very thorough tests have shown that the mechanism provides an accuracy of $\pm 0{\cdot}001$ in under all conditions for the shorter film movements, and never more than $\pm 0{\cdot}003$ in error for the longest movements.

(7) REPRODUCTION OR PLAYBACK

(7.1) General Considerations

Reproduction of the signals recorded on the film is obtained by projecting a fine line of light across the film and measuring with photocells the luminous flux passing through the various tracks. In comparison with other methods this offers the following advantages:

(a) Similarity to standard cinematographic sound reproduction.
(b) Simple associated equipment.
(c) There is no need for any modulating system, at carrier frequency, with its attendant complications.
(d) The output is unaffected by the lateral weave of the film in passing through the mechanism.

The method also has disadvantages:

(a) The output is affected by any imperfection on the film, particularly dirt and fog.
(b) The amplifier is likely to pick up hum at its input.
(c) Errors may be introduced from the lamp, the optical system, the slit which forms the line of light and from image-spread occurring in the optical system and the negative and positive films.

The arrangement, shown in Fig. 11, comprises a lamp with concentrated filament, which is projected by the condenser lenses on to the objective lens. Viewed from the latter lens the condenser would then appear as a uniformly illuminated disc, of luminance equal to that of the filament. A graticule adjacent to the condenser lens is opaque except for a narrow line which is projected on to the film by the objective. Light passing through either of the measuring tracks is reflected on to the appropriate photocell by a mirror behind the film. Each mirror is concave and serves to form an image of the lamp filament (in the objective) on the photo-cathode, with the intention that the patch of light is not altered in shape or position, but only in intensity, as a variable-area track is measured.

In such a system the results are dependent on many factors which, as usual, are often mutually conflicting. For any particular application the choice therefore requires a compromise between them in order to minimize the errors and avoid undue complication of the equipment. It is useful to classify these factors into three groups: those determining the magnitude of the signal derived from the photocells; those affecting the linearity of reproduction; and those affecting the frequency response.

(7.1.1) Magnitude of Signal from Photocells.

To simplify the amplifier it is desirable to obtain the maximum signal from the photocells. The luminous flux on the photocells is dependent in the first place on the power of the lamp; this is limited by considerations of the power supply (which has to be regulated) and heat dissipation, and on its filament which should be of a size and shape to fill the objective lens completely. Secondly, the condenser lenses, whilst powerful, must not suffer such aberrations that the uniformity of luminance across them is impaired. The flux is dependent on the maximum width of each track across the film and hence is inversely proportional to the number of channels recorded; also it is proportional to the dimension, along the film, of the illuminated slit. In practice this is the most important parameter and the only one which may be readily altered. Lastly, in the case of variable-density recording, the flux depends on the mean density (see Section 8.1). The maximum flux is limited only by the permissible illumination of the photocells.

The signal to the amplifier is proportional to the luminous flux, the sensitivity of the photocells (the antimony-caesium vacuum type offer the best compromise of sensitivity, frequency response and stability) and the value of the resistive load associated with each photocell.

(7.1.2) Linearity of Response.

The linearity of response is dependent on the uniformity of luminance of the condenser lenses, on the parallelism of the slit and on second-order errors associated with non-uniform sensitivity of the photocell cathode. The latter turns out to be the most difficult to satisfy; a testing technique, traversing a small spot of light across the scanning position, enabled reasonably good photocells to be selected.

(7.1.3) Frequency Response.

The upper limit of frequency response is primarily determined by the dimension of the slit, in the direction of motion of the film in the same manner as for recording (see Fig. 7); hence, in practice, the compromise has to be made between signal level and frequency response. It is also limited by light spread occurring in the optical system and in the exposure of both the negative and the print, and by the stray capacitances which appear in parallel with the photocell loads.

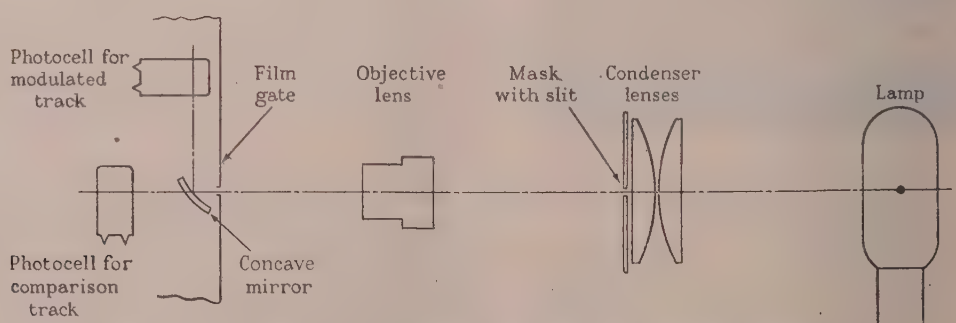


Fig. 11.—Arrangement of optical components used in the reproducing system.

(7.2) Use of Comparison Track

A most important feature for the attainment of accuracy of the amplitude of the reproduced signal, and for reduction of noise, is the recording of a comparison or reference track, which is unmodulated. It is particularly important for the variable-density process.

The signal obtained from the comparison track on reproduction is subtracted from those obtained from the modulated tracks. This gives good cancellation of effects of image spread on variable-area records, and of mains hum and spurious signals picked up in the input circuits of the reproducing amplifier. The alternative process of division would have given better cancellation of errors arising from any imperfections in the film affecting the tracks equally, or from irregularities in film movement during variable-density recording; it would also have fully cancelled fluctuations of the output of the lamp used for reproduction. But the electronic circuits required for this division—or any equivalent process—would be much more complicated than those for subtraction, and the system was therefore not used.

The light flux transmitted through the comparison track is made equal to that transmitted through either of the modulated tracks when at the mid-point of their modulation range. This gives optimum cancellation of all errors lying near the middle of the modulation range, but a progressive increase in the errors in signals of amplitude approaching either maxima. In the majority of applications, and particularly those of a statistical nature, the major proportion of the information can be carried by signals near the mid-point of the range.

(7.3) Exposures for Setting-up and Calibration

Before a recording is made, exposures should be made on the film to permit the reproducing equipment to be set up and calibrated. To set the zero level for reproduction, an exposure is required with zero voltage applied to the input of the recording amplifier; to set the calibration, or scale factor, an exposure is required with a known voltage which gives nearly the full modulation.

This technique is an important factor in the attainment of the accuracy of the whole system, which is made in effect into a comparator. In certain applications it may constitute a limitation, but for the purpose for which the apparatus was designed it was recognized that calibration would be required of the other equipment providing the signals being recorded. With variable-density recording, calibration is needed on every length of film, although this might be relaxed for a group of films of the same batch, exposed and processed together. For the variable-area system, experience may show that calibration at the time of recording is unnecessary, for the errors can be only second-order effects. The reproducing equipment has, however, been designed on the assumption that a calibration film—which could be a separate special length—would be available; the major variables to be corrected are the zero shift of the amplifiers and variation in light output, which affect the zero setting and the sensitivity, respectively.

(7.4) Equipment for Reproduction or Play-back

Even with a lamp with a heavy filament (10 volt, 7·5 amp) and the use of the comparison track, a 100-c/s error signal would appear at the output if the lamp were fed with a 50-c/s alternating current. Accordingly, the mains supply after being transformed to a low voltage is rectified and smoothed to some extent (using a 5-mH choke and a 1 000-μF condenser). A saturated-core regulating transformer is used to minimize mains-voltage fluctuations, although it is necessary to correct for the consequent slow changes in light output caused by changes in supply frequency. Provision is made for running the lamp at half brightness to prolong its life, when it is possible to choose a slit which permits adequate output under this condition.

The electronic circuit is shown in Fig. 12. Each of the three main photocells has associated with it a 3-stage amplifier, in which about 20 dB of negative feedback is used. During manufacture, the load R_1 was matched to each photocell, using a standard test film, to compensate for variations in sensitivity of the photocells; this permits the electronic units to be interchanged on the film mechanisms. A preset control R_6 on each amplifier for a modulated track permits the gain to be made equal to that for the unmodulated track.

The signal from the unmodulated track is subtracted from the signal from each of the other tracks by inverting its phase and adding in the network of R_8.

To prevent any interaction between one measuring-track amplifier to the other, the points marked X in Fig. 12 must present a very low impedance. This is obtained by making V_8 and V_9 into an anode-follower stage, the feedback also improving the linearity.

A requirement of the equipment is that it should give a single-sided, not push-pull output, although push-pull stages were required throughout the amplifier to obtain stability. The circuit between V_8 and V_{11} applies degeneration to hold the junction of the two resistors R_{14} at a constant potential with respect to earth, even if the push-pull stages are out of balance. R_{15} controls the sensitivity of the reproducing equipment, when using the calibrated film, and also permits the phase of the output to be reversed, equivalent to changing the sign in the mathematical analysis.

It is, of course, necessary to maintain a sign relationship throughout the whole recording and reproducing system, the convention adopted being that the application of a positive voltage to the input terminal 1 of the recorder amplifier, with respect to earth or terminal 2, gives a print with increased light transmission, which, in turn, gives a positive potential at the output during reproduction, with R_{15} turned towards V_8.

(7.5) Performance of Reproducing Equipment

The linearity of the reproducing equipment alone was measured by applying known voltages across the input resistor R_1. The output was linear within 0·5% to ±40 volts, the normal designed output being ±25 volts. General considerations and experiments on the photocells used for monitoring the variable-density recording (which are of the same type as those used for reproduction) suggest that non-linearity in the photocells will be less than 1% over the range of illumination used.

The frequency response of the reproducing amplifier, together with the photocells and interconnecting cables, was measured by running a film with a sinusoidal exposure through the mechanism at different speeds. The curve is given in Fig. 13, the rising characteristic being due to the negative feedback in the amplifier of V_{1-3}; it is of value for offsetting to some extent the tendency to h.f. attenuation in other parts of the system.

(7.6) Reproduction from Timing Track

The photocell for reproduction from the timing track is connected to a very simple amplifier, consisting of one push-pull stage of amplification leading to cathode-followers, with no negative feedback. The accuracy of the instant of reading any timing marks on the film is quite high; the slits for exposing and reading the timing track can be adjusted to have very little displacement along the film from the slits for the other tracks; the greatest error is probably associated with the width of the slit on recording, about 0·012 in when projected on the film.

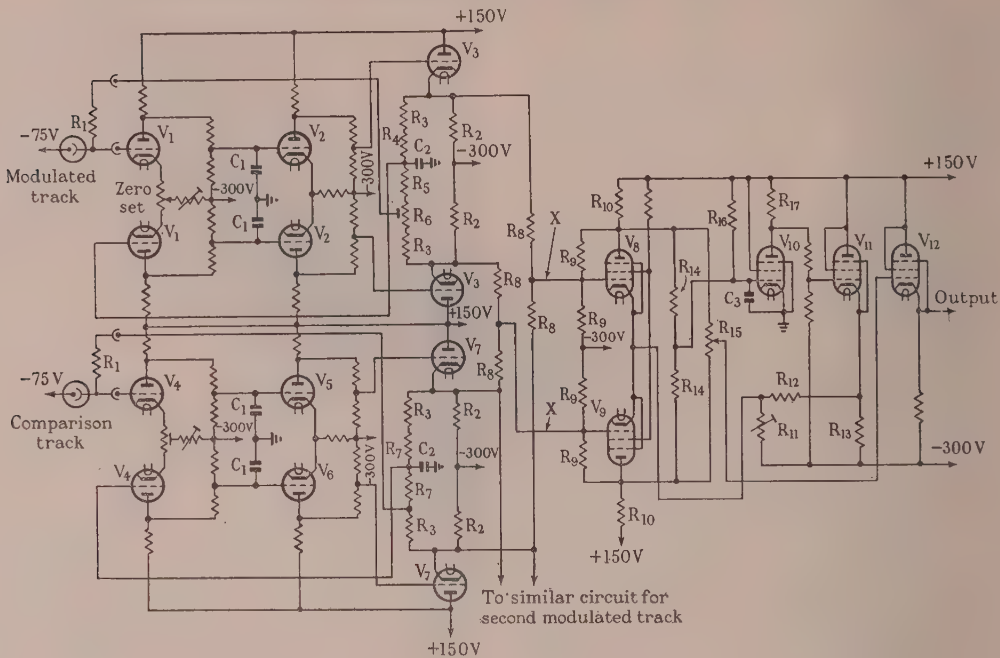


Fig. 12.—Circuit of reproducing amplifier.

V_1–V_7—ECC35.
V_8–V_{12}—EF91.
Photocells—Cintel VA42.
R_1—1 MΩ nominal.
R_2, R_{15}—50 kΩ.
R_3—200 kΩ.
R_4—10 kΩ.
R_5, R_6—5 kΩ.
R_7—8·2 kΩ.
R_8—100 kΩ.
R_9—1 MΩ.
R_{10}, R_{13}, R_{17}—30 kΩ.
R_{11}—20 kΩ.
R_{12}—15 kΩ.
R_{14}—68 kΩ.
R_{16}—1·5 MΩ.
C_1—0·0005 μF.
C_2—0·005 μF.
C_3—0·001 μF.

(7.7) Conclusions on the Design of the Reproducing Equipment

Following the experience of developing this equipment certain alternatives are possible. Firstly, it would probably have been preferable to use larger photocells; these could not be mounted directly behind the film gate, and an optical system would be required, but this would ensure a greater independence from the non-uniformity of sensitivity over the photo-cathodes. Secondly, a more powerful lamp could have been used without overloading the photocells—the present lamp does not entirely fill the objective lens. Thirdly, it would almost certainly have been preferable to mount a cathode-follower valve adjacent to each photocell on the film-mechanism; it was feared that the impairment of the zero stability precluded this, but in fact the zero-drift is unimportant, whilst a certain amount of difficulty has resulted from the use of the high-impedance line throughout the cable connecting the photocells to the amplifier.

(8) PERFORMANCE OF COMPLETE EQUIPMENT

(8.1) Performance of Variable-Density System

The amplitude linearity of the variable-density system is very dependent on the photographic properties of the film. Details of the special processing and the experiments to determine the optimum conditions are given in Appendix 11.4. The requirement is to obtain a linear relationship between the exposing light and the light transmitted through the print. The photographic process is, however, essentially logarithmic, and a plot of the logarithm of the exposure against the density (the logarithm of the reciprocal of the transmission) has a straight portion, its slope being the gamma; the gamma is controllable to some extent

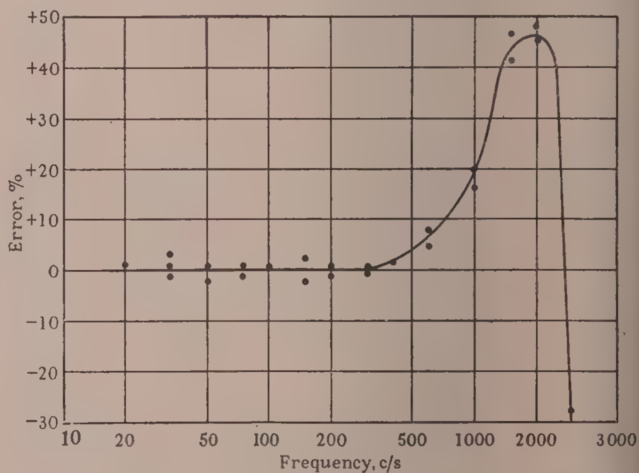


Fig. 13.—Frequency-response characteristic of reproducing photocells and amplifier.

in the processing. This relationship applies, of course, to the exposure and processing of both the negative and the print. In most photographic applications, at the extremes of the range, it is usual to run off the straight part of the characteristic; in this case it is necessary to operate on the curved toe, to obtain sufficient light transmission through the variable-density print.

Optimum conditions therefore amount to an empirical compromise. For good overall linearity the print has to be made

vith only 10% transmission at the mid-point of the modulation ange; to compensate for this, the slit on reproduction is normally ·010in wide at the film, with attendant impairment of the igher-frequency signals. This may constitute a significant imitation of the variable-density system, unless an increase of oise level from the amplifier can be tolerated.

The overall linearity was determined by applying voltages to he recorder from a calibrated switch-box, and measuring the utput from the reproducing amplifier with a sub-standard volt-neter (1 000 ohms/volt). The film was moved slowly and a mean eading taken, to average out the irregularities in the film. Fig. 3(*b*) gives the errors, which fall within ±1·5% except at he extremes of the range of modulation.

To show the noise level a recording was made (at 5in/sec) vhile the voltage at the input was changed suddenly from 0 to ·1 volt on the ±1 volt range. The output on reproduction at 1in/sec) was displayed on a cathode-ray oscillograph photo-raphed with a moving-film camera, as in Fig. 14(*a*), showing he noise to be about 3% peak-to-peak.

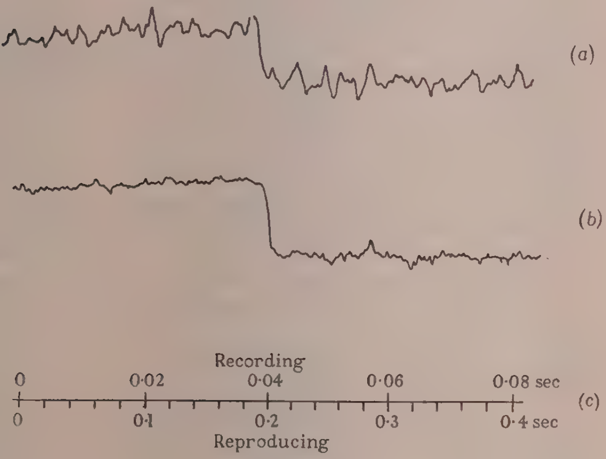


Fig. 14.—Noise levels of complete equipment.

Traces derived from a cathode-ray oscillograph connected to the output. The libration step represents 5% of the full range of recorded signal.

(*a*) Variable-density system.
(*b*) Variable-area system.
(*c*) Time scales.

To determine the frequency response of the variable-density lm, a recording was made with signals of constant amplitude, nd on reproduction the output was measured by the backing-off chnique. The results are shown in Fig. 4(*b*). In such a system, ttenuation at high frequencies (corresponding to short wave-ngths) would be expected owing to the sizes of the slits used r recording and reproducing (0·0025in), and light scatter in the optical systems and films; however, this is seen to have been offset somewhat by the rising frequency characteristic of the recording amplifier [Fig. 4(*a*)]. In this experiment the film was recorded at 10in/sec and reproduced at 2in/sec; the results are therefore not affected by the rising frequency characteristic of the reproducing amplifier itself (Fig. 13).

(8.2) Performance of Variable-Area System

Corresponding tests were made to determine the overall performance of the variable-area system. In general, these were carried out with a greater precision than the variable-density tests, partly owing to the better performance given by the system, and partly as a result of improvements in the experimental techniques.

The amplitude linearity is shown in Fig. 3(*d*); the results are slightly worse than those shown in Fig. 3(*c*), which is due to the effect of non-uniformity of the sensitivity over the cathodes of the photocells, although this is minimized by the design of the optical system and the integrating action along the slit.

Fig. 14(*b*), derived in the same way as Fig. 14(*a*), shows that the noise level seldom exceeds 1% peak-to-peak; at the maximum amplitude (widest track) this increases to 1·5%.

The frequency response of the variable-area system was found to be affected by light-scatter in the optical systems and the films; at the boundary between areas of great and small light intensity, some of the light is scattered, causing either an increase of the effective width of the slits or a certain amount of exposure of the immediately adjoining area of the film. Some compensation for these effects was made by the choice of the exposures given to the negative and the print, but the other parameters—the grades of film and the processing conditions—were not varied in this application, standard cinematographic practice being adopted.

Experiments showed that there is an attenuation of the reproduced signal which is solely dependent on the wavelength of the signal on the film; its magnitude is given in Fig. 9(*b*), and it is probable that there are some applications where this would constitute the major limitation of the equipment. For this test the film was recorded at 10in/sec and the positive and negative peak amplitudes of the signal applied to the recording amplifier were measured with the backing-off technique, and kept constant; on reproduction, the output was measured in the same manner, but the film speed was reduced, so as to use the level part of the characteristic of the reproducing amplifier (Fig. 13). The width of the slit was 0·0013in.

The attenuations of the positive and negative peak amplitudes are not quite the same, this being equivalent to the introduction of some harmonic distortion [Fig. 9(*b*)].

It is believed that this amount of attenuation at short wavelengths is of the same order as that usual in sound-recording equipments, provided that allowance is made for the essential differences between the systems; probably the most important

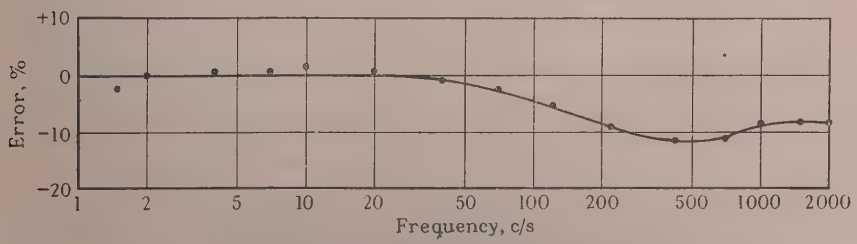


Fig. 15.—Frequency response of variable-area system, overall.

Film speed on recording and reproduction: 10in/sec.

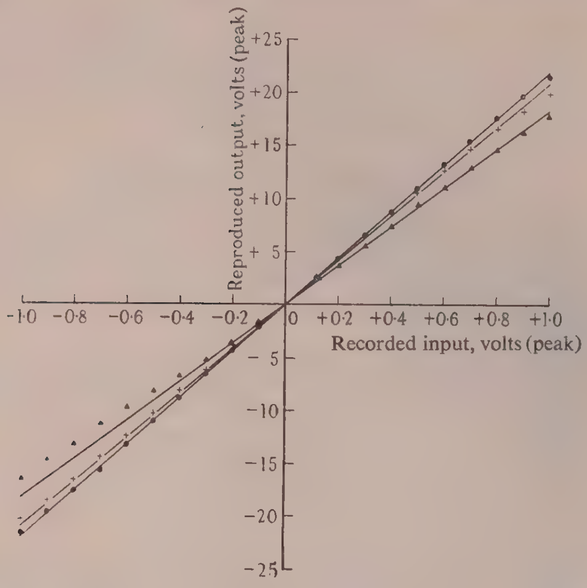


Fig. 16.—Direct plots to show the amplitude linearity of the variable-area system (overall) for sinusoidal signals at three frequencies.

● 10 c/s.
+ 120 c/s.
▲ 1 000 c/s.

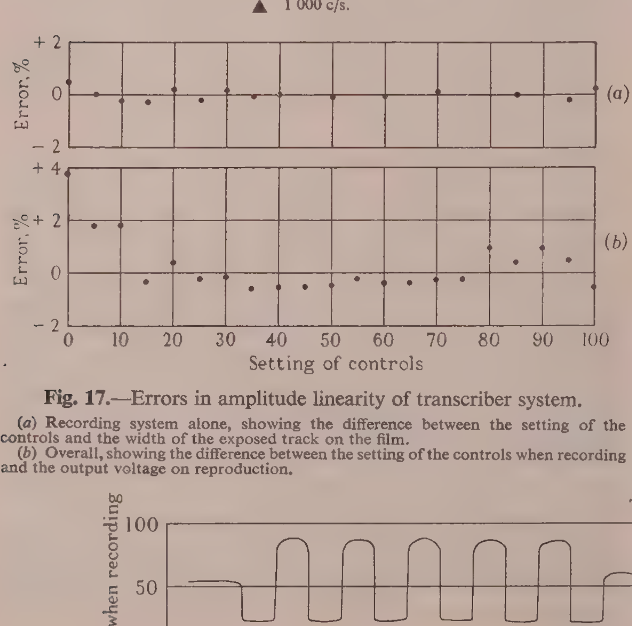


Fig. 17.—Errors in amplitude linearity of transcriber system.

(*a*) Recording system alone, showing the difference between the setting of the controls and the width of the exposed track on the film.
(*b*) Overall, showing the difference between the setting of the controls when recording and the output voltage on reproduction.

point is that the present optical system was designed to cover the whole width of the film (to permit the recording of several tracks), with consequent increase of throw from the objective lens to the film.

When the film of Fig. 9(*b*) was passed through the reproducing equipment at 10 in/sec the results of Fig. 15 were obtained. Some compensation is obtained from the rising frequency characteristic of the reproducing amplifier, and it is seen that under the best conditions the frequency response of the whole system has a maximum error of about 10%.

The specification of the overall frequency response is consequently somewhat complex, owing fundamentally to the facility for being able to record at a range of film speeds and being able to reproduce the film at a different speed from that used when recording. For certain applications it might be preferable to introduce some form of rising frequency characteristic in the recording process; but this was not done in the present application since the primary requirement was good amplitude accuracy at low frequencies.

To determine whether the attenuation at short wavelengths is in any way related to the amplitude of the signal, a sinusoidal wave of frequency 10 c/s was recorded at ten different amplitudes up to the maximum; this was repeated at 120 c/s and 1 000 c/s. The film speed on recording was 10 in/sec, but was slow on reproduction; the backing-off technique was used to measure the peak amplitudes on recording and reproducing. A direct plot of the results is given in Fig. 16, showing that

(*a*) The amplitude accuracy for alternating signals is good.
(*b*) There is a slight attenuation at 120 c/s, becoming significant at 1 000 c/s.
(*c*) The attenuation of negative peaks at 1 000 c/s is consistently greater than that of the positive peaks.

Throughout all the tests, no evidence was found of cross-modulation between the various tracks on the film.

(8.3) Performance of Transcriber

The amplitude accuracy of the transcriber recording system alone was determined by measuring the width of the exposed tracks with a travelling microscope, the errors being given in Fig. 17(*a*). This shows that the arrangement of moving masks is satisfactory and that the optical system introduces little distortion.

Fig. 17(*b*) shows the amplitude errors of the whole transcriber system, including the reproducing equipment. Each exposed track is modulated on one side only, which somewhat exaggerates

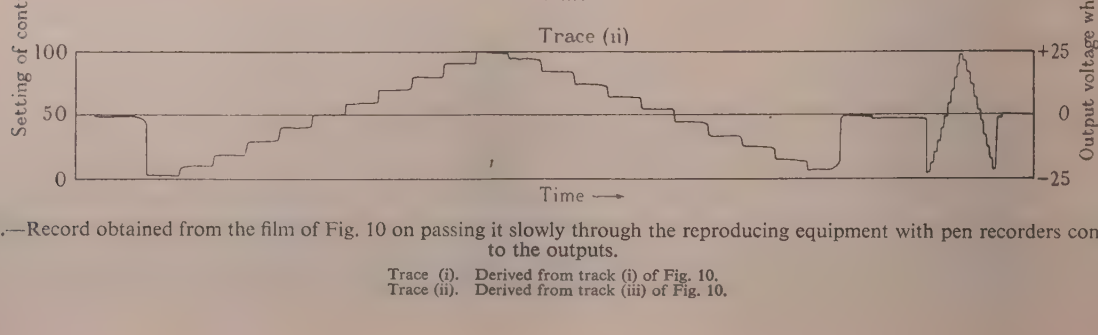


Fig. 18.—Record obtained from the film of Fig. 10 on passing it slowly through the reproducing equipment with pen recorders connected to the outputs.

Trace (i). Derived from track (i) of Fig. 10.
Trace (ii). Derived from track (iii) of Fig. 10.

the errors due to the non-uniformity of the sensitivity of the photocells.

Fig. 18 is based on a record obtained with a high-speed, twin-pen chart recorder, when the film of Fig. 10 was passed very slowly through the reproducing equipment. It will be noticed that the effect of light-spread results in a slight rounding of the more positive parts of the square wave at $\frac{1}{10}$in film movement, and a reduction of the amplitude of the corresponding parts at $\frac{1}{100}$in movement. Such errors would be less severe if the difference in amplitude between successive exposures were smaller, as would usually be the case.

(9) ACKNOWLEDGMENTS

The development of this equipment was based on the previous work of Messrs. H. W. Pout and K. W. Thwaites of the Admiralty Signal and Radar Establishment and continued with their co-operation, and that of others. The staff of Cathanode Ltd. completed the detailed design and manufacture of the film mechanisms; Mr. A. S. Pratt of British Acoustic Films Ltd. advised on the film transport, and designed and adapted the variable-area light modulators; Mr. W. M. Wreathall of Taylor, Taylor and Hobson Ltd. designed the optical systems; Messrs. F. P. Gloyns and J. Luscombe of Denham Laboratories Ltd. evolved the precision photographic processing methods; Mr. J. E. Pateman suggested many features of the electronic circuits; Mr. P. Briggs gave the circuit for the servo-motor drive of the film mechanisms; Mr. F. Rock-Carling suggested the pawl and ratchet arrangement for the transcriber; and Messrs. G. E. Felton, P. Flitter, L. R. Forryan and J. G. C. Edwards also assisted.

Acknowledgment is made to Elliott Brothers (London) Ltd. and to the Admiralty for permission to publish this paper.

(10) REFERENCES

(1) Darbyshire, J. A.: "A New Modulated Light Source," *Electronic Engineering*, 1951, **23**, p. 167.

(2) Mees, C. E. K.: "The Theory of the Photographic Process," (Macmillan, New York, 1942).

(3) Frayne, J. G., and Wolfe, H.: "Sound Recording" (Wiley, New York, 1949).

(4) Endall, R.: "Magnetic Tape Data-Recording Systems," *Radio-Electronic Engineering*, March, 1952, **30**, p. 3.

(5) Boothe, K. B.: "Tape Recording for Telemetry and Data Analysis," *Tele-Tech*, May, 1952, **11**, p. 44, and June, 1952, p. 56.

(6) Begun, S. J.: "Magnetic Recording" (Murray Hill, New York, 1949).

(7) Skellett, A. M., Leveridge, L. E., and Gratian, J. W.: "Electron-Beam Head for Magnetic Tape Playback," *Electronics*, October, 1953, **26**, p. 168.

(8) "Multi-Channel Magnetic Tape Recorder for Slowly Varying Direct Voltages" (Elliott Bros. Ltd.), *Handbook of Scientific Instruments and Apparatus* (Physical Society, London, 1953), p. 248.

(9) Davies, H.: "The Design of a High-Fidelity Disc Recording Equipment," *Journal I.E.E.*, 1947, **94**, Part III, p. 275.

(10) Williams, F. E.: "The Design of a Balanced-Armature Cutter-Head for Lateral-Cut Disc Recording," *Proceedings I.E.E.*, Paper No. 805 R, March, 1949 (**96**, Part III, p. 145).

(11) APPENDICES

(11.1) Typical Analyses of Recorded Data

It is visualized that typical forms of analysis of the data recorded and reproduced with this equipment will include the following:

(*a*) *Computation of Means.*

$$\bar{f} = \frac{1}{T}\int_0^T f(t)dt$$

(*b*) *Computation of Mean Squares.*

$$\overline{f^2} = \frac{1}{T}\int_0^T f(t)^2 dt$$

(*c*) *Computation of Variances.*

$$\sigma^2 = \frac{1}{T}\int_0^T [f(t) - \bar{f}]^2 dt$$

(*d*) *Fourier Analysis or Power Spectra.*

$$a_n = \frac{2}{T}\int_0^T [f(t) - \bar{f}] \begin{matrix}\sin\\ \cos\end{matrix} \left(\frac{2\pi nt}{T}\right) dt$$

The sinusoidal voltage required for this computation may be generated with a sine-cos potentiometer, for example, or by running a film carrying a sine wave through a second reproducing equipment; it is expected that that film might conveniently be recorded on the transcriber. This method of determining power spectra may be more convenient than deriving them from the auto-correlation functions.

(*e*) *Cross-correlation coefficients*, generating simultaneously two functions $f(t)$ and $g(t)$ from two reproducing equipments.

$$\rho = \frac{1}{\sigma_f \sigma_g}\int_0^T [f(t) - \bar{f}][g(t) - \bar{g}]dt$$

(*f*) *Analysis of line-of-sight experiments*, where the co-ordinates of the object in the field of view (recorded with a cine-theodolite) are transferred by means of the transcriber and the reproducing equipment to the computer.

(*g*) *Input of the forcing function* to a differential analyser or to an analogue computer set up for the solution of single or simultaneous differential equations; and the introduction of non-linear relationships into an analogue computer.

(11.2) Features of Methods of Recording and Reproduction

For applications of the types outlined in Section 1, the features of the various techniques for recording and reproduction may be set out as follows. In general, the optimum performance in any one respect can be attained only by impairment of the performance in other respects.

(11.2.1) Cinematographic System.

Digital or analogue methods may be used. In the digital system each word may be recorded across the width of the film, many digits being accommodated with consequent high accuracy. With analogue methods the amplitude accuracy may be ± 1 or $\pm 2\%$ of full scale.

With analogue methods the frequency range extends from zero to perhaps 10 kc/s. The electro-mechanical nature of the light-modulating element introduces a limitation.

The noise level is approximately 1% (−40 dB) of maximum signal, measured over the audio-frequency band.

The time scale may be changed over a ratio of 100 : 1 with ease, and of 1 000 : 1 without difficulty. An intermittent movement may readily be obtained.

The accuracy of the time scale is good over long periods, the extreme accuracy of the perforations contributing towards this.

Short-period irregularities are usually 1 to 0·1%, partly owing to mechanical imperfections and partly arising from effects associated with the perforations.

The equipment for analogue recording is simple.

Processing of the records is required. If the delay can be accepted it is most desirable to transfer this work to a specialist laboratory. Rapid processing may be carried out, but with considerable complication.

No facility for immediate playback is available.

A darkroom is required for loading film into the cassettes, but these may usually be put into the recorder in daylight.

Any failure of the take-up mechanism within the recorder can be serious since it is not readily observed.

Film may only be used once in recording, and material costs may be significant in some applications.

The records are permanent. Duplicate prints may be obtained, and the primary record (the negative) is safe and need only be handled within the processing laboratory.

No erasure of mistakes from the records is possible except by editing and jointing.

Multiple channels may be recorded.

The storage capacity, in terms of quantity of recorded information per unit volume, is relatively small unless multiple channels have been recorded.

There is no rapid access to any particular point on the record, which is wound up in a reel, but the point can be readily identified by the footage number which is visible on the negative before and after processing and is printed on the final film; the record can be studied visually.

The cinematographic industry is relatively old, and the technical processes are well established.

(11.2.2) **Magnetic Tape System.**

Immediate playback may be obtained.

Tape may be used repeatedly, with consequent saving in cost in some applications. Mistakes can be erased and corrections made.

Digital or analogue methods of recording may be used.

The usual pick-up device is responsive to the rate of change of magnetic flux and consequently has no d.c. response. Some form of modulation is therefore required. With an analogue system and an amplitude-modulated carrier the errors may amount to 10% unless selected tape is used.

With frequency modulation, overall accuracies of $\pm 1\%$ may be obtained.

A typical upper limit of frequency response is 10 kc/s, but higher frequencies have been obtained, using a carrier of 100 kc/s.

It is not normally practical to obtain a change of time scale greater than about 10 : 1.

The short-period constancy of speed of tape movement is usually about $\pm 0{\cdot}5\%$, but improved equipment is becoming available with uniformity within 0·1%. Short-period irregularities appear as noise in the output.

Apart from noise introduced by speed irregularities, the signal/noise ratio is very high, for example 50 dB over the audio-frequency band.

Multiple channels may be recorded, either side by side on wide tape or on one track with a sequential sampling system.

The storage capacity is great.

There is no rapid access to any particular point in the record.

The record cannot normally be interpreted visually; localized inspection may be obtained by applying iron powder.

The original record is normally used for playing back.

The use of the technique is relatively novel, and it is reasonable to expect that progressive improvement will occur.

(11.2.3) **Disc Recording.**

Extremely rapid access is given to any point on the record.

Immediate playback may be obtained.

The storage capacity is very great.

The pick-up device has no d.c. response. Mechanical resonances limit the frequency response to the a.f. range, and within at least the lower part of this range the amplitude error is probably less than 5%. The accuracy might be increased by amplitude or frequency modulation of a carrier, with consequent extension of the frequency range to zero but with restriction of the upper limit. The recorder is affected by external vibration.

The constancy of speed is determined solely by the mechanical drive to the turntable and centring of the disc; short-period irregularities may be made less than 0·2%.

There is a gradual change of conditions across the disc, chiefly affecting the upper-frequency response.

The signal/noise ratio is good, being about 55 dB over the audio-frequency band.

(11.3) Circuit of Variable-Density Recording Amplifier

The circuit for each modulated variable-density recording lamp is given in Fig. 19. The V_1 stage, being outside the feedback loop, is made linear by R_3 and gives little gain. Its output is added by R_7 and R_8 to the output from a similar stage, V_2, associated with the monitoring photocell.

An input attenuator provides for full modulation for ± 1, ± 2, ± 5, ± 10, ± 20 or ± 50 volts, applied differentially between input terminals 1 and 2. The circuit is, however, largely independent to voltages applied in the same polarity to both terminals.

The gain of the amplifier is adjusted by selecting by trial an appropriate value for R_1, the load for the photocell; this compensates for differences of sensitivity of various photocells and determines the amount of feedback. The criterion is that, with a typical lamp, the application of ± 1 volt at the input should swing the lamp current over the range 5–120 mA.

The circuit for the unmodulated lamp is similar to V_3–V_5; since there is no requirement to inject modulating voltages, V_1 is dispensed with and the photocell is connected directly to V_3. No cathode-ray tube is required.

The power-supply points are connected to one neon-stabilized pack which is common for all lamps. In addition, each of three independent packs provides the main current for one lamp, being connected to points A and B. These are not stabilized, and in common with all supplies throughout the equipment, use metal rectifiers. The potentials for the cathode-ray tube are derived from the supply which serves the valves.

(11.4) Photographic Conditions for Variable-Density System

For this application the following technique was evolved for processing the variable-density films with the highest possible precision. After exposure of the negative (on Kodak 5231 stock) it is submitted with a length of unexposed film from the same reel. This is exposed in the laboratories in the Type 2B sensitometer and a short length put through the picture-negative processing machine. After processing, the densities of the test strip are measured and plotted (log E against D), the slope of the best straight line being the gamma. Repeated trials are made, with differing times of development, until the gamma is brought to 0·70, it being usually possible to obtain a precision within $\pm 0{\cdot}01$.

The print is made on Kodak 5302 stock and processed in a low-contrast developer, trial runs being made as before to find the processing time required for a gamma of 1·70. The exposure of the print is chosen to give a density of 1·0 on the unmodulated track on the film.

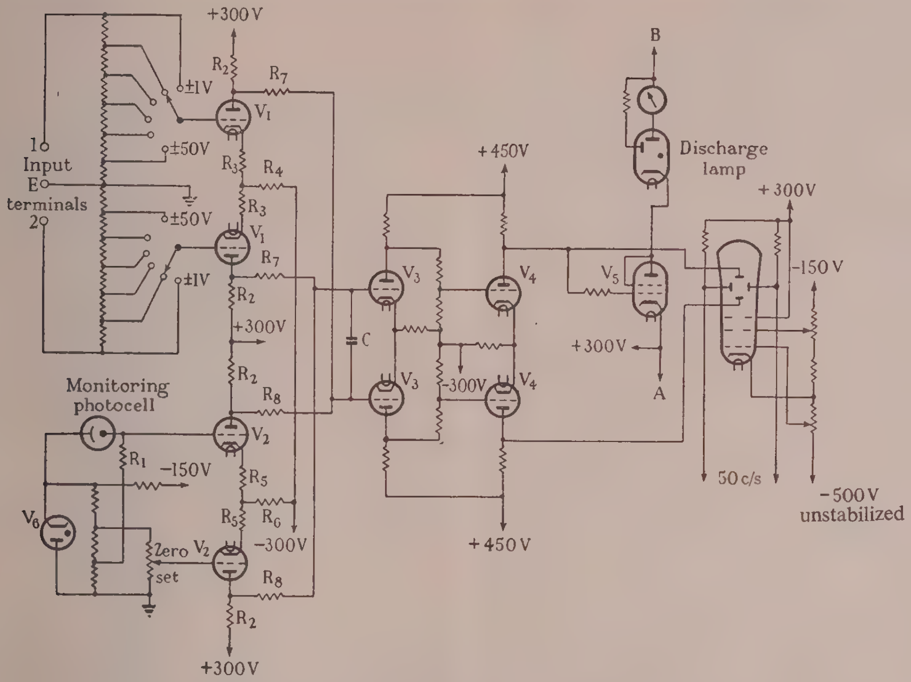


Fig. 19.—Circuit of variable-density recording amplifier.

Points A and B are connected to a separate power supply of 200 volts.

V_1–V_4—ECC35.	R_1—1 MΩ (nominal).
V_5—12E1.	R_2, R_6, R_7—200 kΩ.
V_6—85A2.	R_3, R_8—100 kΩ.
Photocell—Cintel VA42.	R_4—175 kΩ.
Lamp—Ferranti GMC6.	R_5—50 kΩ.
Cathode-ray tube—Mullard DP4/2.	C—0·01 μF.

It will be appreciated that several days must be allowed for carrying through such a procedure, which demands the greatest skill from the laboratory staff. In general, the precisions of the measurements involved are approaching the limits obtainable from the instruments, and are tending to be masked by the normal irregularities of the properties of photographic materials.

All measurements of density (and hence gamma) are appropriate to the instruments at the laboratory; they do not therefore relate to the apparent densities if these could be measured on the recording and reproducing equipment.

Experiments showed that the negative and positive gammas specified gave much the same results as gammas of about 1·0 for both; they were preferred as being nearer to standard practice.

To determine the optimum photographic conditions a series of prints were made from one negative, changing both the gamma and the density (for a gamma of 2·5 the normal high-contrast developer was used). A representative group of results is given in Fig. 20. It shows that the linearity is slightly affected by the gamma and seriously impaired by reduction of density. A further test, with 0·5 density for the unmodulated track, showed marked non-linearity, clearly visible even in a direct plot. The low density is, of course, sought after with the aim of passing more light to the photocells. The scatter of the points is not greater than was expected from a photographic experiment done at an early stage of the equipment.

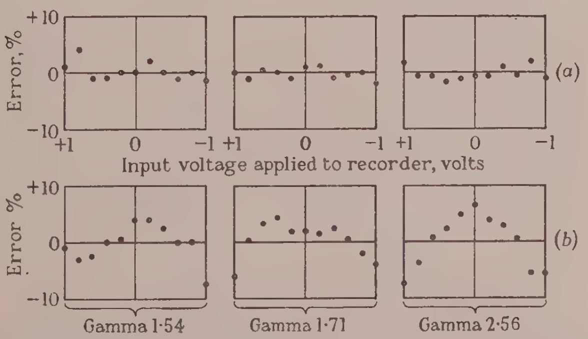


Fig. 20.—Effect of different photographic conditions for the print on the amplitude linearity of the variable-density system, overall.

(a) Density of unmodulated track, 1·0.
(b) Density of unmodulated track, 0·7.

DISCUSSION BEFORE THE MEASUREMENTS AND RADIO SECTIONS, 4TH JANUARY, 1955

Mr. J. F. Coales: The recording of trials data is one of the most difficult of all forms of data processing and handling. It is a problem that has been with us in an extreme degree since before the war. Very large numbers of people have been involved in taking records, particularly, of course, of the movements of aircraft, both in America and in this country, and during actual air-raids very large quantities of information were recorded. This involves many millions of feet of film altogether, and in most cases it has never been analysed because of the enormous amount of labour which is involved.

The author and some of his colleagues were very much concerned with this recording of data, and some of his colleagues were involved in a digital method of recording in which the most complete series of records were compiled in this country, but although in the case of digital recording it was done with a view to using a digital computer for most of the analysis, no complete analysis has in fact been possible owing to the amount of labour required.

It was, of course, in order to solve this problem that the work was begun, and it will be appreciated from the paper that a very considerable step forward has been made in the work of recording and processing trials information. That is only one of the aspects of this type of equipment. Several years ago I was of the opinion that analogue computing was becoming obsolete, and that by about 1955 all computing would be done by digital methods. However, I was engaged in discussions on the analogue computer, Tridac, and we found that, because computations had to be carried out in what we call real time, i.e. on the actual time scale in which the events happen in nature, they could not be done by digital methods, which were not fast enough. If we have a very large number of equations to solve and we have a fixed time scale, with practicable digit repetition rates it is not possible to carry out the computations. For this reason there has been a very real resurgence of analogue-computing methods in the aircraft and other industries.

Analogue computing has other advantages. If a problem is set up on a differential analyser it can quickly be seen how things are going; for example, if a control system is set up, one can see by the behaviour of the integrators, how changing some of the parameters will affect the answer. This is a great advantage, and it cannot easily be done with a digital computer.

I am interested in an on-off control system in which the actual time of change-over of the output motor from full acceleration in one direction—in order to catch up the input—to full deceleration to bring the output into line with the input in both position and velocity has to be computed. One way of doing this is to use a simulator of the output system operating 100 times more rapidly than the actual motor and load in order to give information on what is going to happen if the change-over is made at a certain time. If we have a complicated motor and load system, it is, of course, difficult to compute what it will do, but most motors and loads can easily be simulated, and this provides an easy method of doing it; however, the system still involves some analogue computing. The means for increasing speed which is described in the paper is not directly applicable in this case, but it is certainly of interest. I think that there will be ever-increasing uses for this type of analogue computing in control systems and in automatic control in general.

In order to simplify automatic control, I am confident that we shall turn more and more to non-linear controls, particularly on-off controls which will become popular on account of their simplicity. This involves some method of assessing their performance for random inputs. As the author stated, this cannot in general be done satisfactorily using step functions and sinusoidal inputs. We have a great deal of investigation to do both of the types of fluctuation that we get at the inputs to our systems and of noise that may arise within the system, and we want methods of recording typical input data, of analysing these and of getting correlation functions and, in the non-linear case, probability distributions. This type of equipment undoubtedly has a future in this respect.

I am concerned about the frequency response. For most automatic control problems 1 000 c/s is probably adequate, but in the computing field I do not think that it is. In the application to on-off controls, there was some need to go to repetition rates of more than 100 per second, which might mean that we should want to have response at higher frequencies, say about 10 kc/s; I wonder whether there is any possibility of improvement in that direction.

The author seems to contradict himself somewhat on the question of subtraction or division; he seems to indicate that division would be better but more complicated.

I agree with the author on the question of recording datum and scale factor. Anyone who takes miles of film or magnetic tape and does not insert the datum and scale factor when he starts the run is heading for trouble.

Is it possible to transcribe on to a film which has already one track exposed with suitable correlation between the tracks?

Mr. K. W. Thwaites: We have found in practice that a very real advantage is gained by being able to look at a specimen of the record before it is sent away for processing. This applies particularly to the variable-area system, in which, after taking a record of a particular operation, one is able to take away a small piece of film, perhaps containing a calibration strip, develop it —not very precisely because it is not critical—and then measure the width of the tracks with a travelling microscope. This gives an indication of whether the record will be of any use, and gives one more confidence in sending away the record for processing.

After the record has been processed, it is very useful to be able to identify specific phenomena—not only those which one expects to find but those which are suspected and about which there is no certainty. For example, the output of some radar device may be recorded, and during a run, a fault may develop which results in oscillation. If the data were run through a reader, and the output fed into the computer, with no examination of the record, we would be unaware that there was anything peculiar about it and a spurious answer might result. If the film were first run through the rewinder the fault would become apparent.

We go further than Mr. Coales and put calibrations on films of the particular variable which we are recording. For example, if we were recording the angular movement in two co-ordinates of a radar beam, we should not only record the calibration of the instrument in terms of voltage, but we should also insert a calibration strip giving the angle of beam motion. Therefore, when the film has been processed it gives all the information required for complete analysis.

A great deal of trouble can often be saved if one can see the data as they are being recorded. Therefore we have made a standard practice of connecting in parallel with this device either a pen recorder of the conventional type or an oscillograph.

It often happens that we require to determine the correlation between two simultaneously occurring functions. Both variables can be recorded simultaneously on the two tracks, and they can be replayed on the reader and fed directly into a computer. An example of this is the determination of the amount of cross-coupling in two orthogonal servo-control systems in which the resolving system is not perfect and we are not certain how imperfect it is. The fact that it is a d.c. machine and has constant-amplitude response over a considerable band of frequencies is also very useful, because one has an opportunity of applying any particular smoothing process desired on the data during the reading process and not during the recording process. This leads to a wider range of treatment which can be applied to the data.

One particular application was, I think, not intended. The upper frequency response of the recorder and reader is limited partly by the finite width of the slit. This provides a ready-made machine which applies rectangular smoothing to data, and I propose to use this property for comparing the efficiencies of practical filters, on a particular type of data, with the theoretical rectangular filter.

Another application of the transcriber is for the recording of standard waveforms on to films for test purposes. One of the most useful films we have at present is of random numbers. The frequency-shift ability of the system is also very useful because we can take the response of relatively slow-moving devices such as mechanical servo mechanisms, speed up the film during reading and determine the power spectra with a conventional spectrum analyser.

Mr. W. Bamford: It often happens that random transient occurrences require to be recorded—an example being a fault on a power-supply network. To do this in the traditional way, i.e. by continuous recording, is very expensive even with low chart speeds, and unsatisfactory because of the very short duration of the occurrence and the cramped record.

One instrument is in use which meets these disadvantages by printing a record only for a short time which, by a simple and ingenious method, includes the few seconds before the fault occurs.

Another device has been used which includes a continuously running loop on a magnetic-tape recorder. By spacing the recording and the pick-up heads at different points on the loop a "memory" can be created, in time equal to the time taken for the tape to pass from one head to the other. Such a loop can be played back on a cathode-ray tube after the occurrence of a fault.

It is difficult, however, to use it for re-recording permanently because of the mechanical problem of removing a magnetized loop from one machine to another and re-recording accurately, and also because of the complication of extending the time-base in order to improve the legibility of the record. As each loop, on the occurrence of any fault, would automatically cease to run in order to preserve the record, there would have to be a system of collecting the tapes and replacing them in order to provide a reasonably continuous service.

It is now suggested that a centrally disposed film recorder could deal with a number of such tape-recording machines, being connected to any one, as required, through the telecommunication network. On the occurrence of abnormal conditions the pick-up head concerned would transmit a code signal which would couple it to the film recorder and provide a recognition signal. The film would then record from a period a few seconds prior to the fault the signals recorded on the tape until a sufficient time had elapsed. It would automatically expand the time-base sufficiently and provide a permanent record with all data upon it. It seems possible that one film equipment could meet the requirements of several strategically placed magnetic tapes.

I have no conception of the economics of this proposal, nor of its technical possibility. It would, advantageously, require some seven records on one film, plotting three voltage, three current and one time measurements. Could this be done by using a wide film in order not to sacrifice too much of the accuracy which the author claims?

Mr. J. A. Colls (*communicated*): The author rejects all methods employing cathode-ray tubes on the grounds that the lateral weave of the film is inevitable, that deflectional linearity is difficult to achieve, and that complex electronic circuits are required. On the contrary, the first of these objections is easily removed by suitable design of the film gate, while modern cathode-ray tubes with flat screens are more than adequately linear, and any apparent complexity of circuit design should be offset against the mechanical complexity and high power consumption of alternative equipment.

Two types of cathode-ray reproducer may be considered here. One is a line-follower operating on the principles briefly described in Section 1.4, and the other is an edge-follower working on a single-sided variable-area record. Both have the advantage of a feedback loop which includes the film record, and provided that suitable precautions are taken to obviate the effects of burns or other irregularities on the tube screen, the accuracy depends mainly on the deflection characteristics of the tube. The latter in turn are dependent only on the geometry of the tube and the applied anode voltage, which may readily be stabilized to the required accuracy. The system described in the paper, on the other hand, is an open-loop system, in which the linearity and constancy depend on a number of factors such as the light output from the lamp, cleanliness and parallelism of the slit, linearity of the photocell, and amplifier gain, all of which are somewhat difficult to control.

Both systems also have the advantage that they can operate on the original negative record, thus cutting down the amount of photographic processing by a considerable factor, in all cases where multiple copies of the record are not required.

The line-follower has the great advantage that it can accept records made on a wide variety of cathode-ray or Duddell oscillographs, or even ink records on paper, and is thus in many ways the ideal general-purpose reproducer. It is admittedly somewhat complex, but complexity in the reproducer would appear to be relatively unimportant compared with the advantage of simplicity and universal availability of suitable recording equipment. Its performance is also somewhat better than the figures claimed for the variable-area system, the frequency limit being about 100 c/s for 1% error, with good linearity and zero stability.

The edge-follower is very much simpler in circuit design, and can reproduce frequencies up to at least 3 000 c/s, but, of course, it has the disadvantage of requiring special recording equipment. Up to the present, little attention appears to have been given to the development of a suitable cathode-ray recorder for this purpose, and the Duddell type would appear to be simpler and less bulky.

The system described in the paper does not by any means exhaust the possibilities, and the relative merits of other systems should be taken into consideration, especially if the object is to select equipment of general application. One very interesting possibility would be to apply spot-wobble to a cathode-ray recorder, thus producing a broad line, and reproduce by means of an edge-follower working on one edge of the line. Such a system might well be found to combine optimum performance with maximum simplicity.

THE AUTHOR'S REPLY TO THE ABOVE DISCUSSION

Mr. H. McG. Ross (*in reply*): It is very interesting to learn from Mr. Thwaites of the practical uses (some of them unexpected) of this equipment. Having carried out the development work and laboratory testing, it is gratifying to note the ever-increasing applications of the apparatus.

It would probably be possible to meet Mr. Coales's requirement and obtain a useful record at 10kc/s. On the debit side there would be a significant increase in film consumption since it would have to be moved faster, and a general reduction of accuracies arising, in part, from a reduction of damping in the oscillograph unit.

I would not attempt to transcribe a second record on to a film already carrying one trace; it would be far better to record two separate negatives, and then have them combined when the print was made at the film laboratory.

Mr. Bamford's suggestion provides an interesting example of

the combination of magnetic-tape and film-recording techniques. It takes advantage of the particular assets of each, i.e. using the tape over and over again and obtaining the permanent record on the film, which can subsequently be analysed thoroughly.

The use of wider film is to be deprecated; in fact, the quality of the recorded traces with this equipment is very good, and it might well be possible to carry more tracks on the 35 mm film. I feel that in such work electrical engineers should only depart from standard motion-picture cinematographic practices when there are compelling reasons. In the cinematograph industry there is a great fund of knowledge for overcoming all those small practical difficulties which do not appear in reports or textbooks.

I look forward to reading the full report on the equipment proposed by Mr. Colls, particularly with respect to knowing what is meant by its "good linearity and zero stability," and whether its flexibility for reading from many types of record makes it a "jack-of-all-trades" equipment.

All systems using cathode-ray tubes, including those suggested by Mr. Colls, are likely to suffer from the disadvantage that records made with one tube may give erroneous readings when reproduced with another tube; this is due to the local irregularities of the screen and variations in the tube geometry. It is very important that records should not lose their accuracy and validity with the passage of time, nor be affected by replacement of items like tubes or valves.

Unless its effects had been cancelled out, the film weave usual in cinematographic equipment would have used up the whole of the 1% error which was the design target for the present equipment. It is contrary to good cinematographic practice to rely on precise lateral location of traces on the film (for a multiplicity of reasons), and this principle should only be departed from when precise tests have shown the effect to be unimportant in any particular application.

The disadvantages of the open-loop system of the present equipment are recognized, although they are lessened by being relevant only during playback, which is done in the laboratory and may be repeated. All alternative systems (including magnetic tape) seem to require, during recording, more complex equipment or operation. The present variable-area apparatus has—apart from switches—only one adjustable control which has to be correctly set for recording, i.e. the zero-adjuster, and it has been found that the range for this can be so limited that even an extreme error in setting can be cancelled out by the zero-signal calibrating exposure. It is general experience that, at some time or another in the stress of trials recording, every adjustable control on the equipment will be turned to the *wrong* setting.

621.385.1.004.1

Paper No. 1740 R
Dec. 1954

THERMIONIC VALVES OF IMPROVED QUALITY FOR GOVERNMENT AND INDUSTRIAL PURPOSES

By E. G. ROWE, M.Sc., Member, P. WELCH and W. W. WRIGHT, B.Sc.

(*The paper was first received* 2*nd December*, 1953, *in revised form* 14*th April, and in final form* 20*th July*, 1954. *It was published in December*, 1954, *and was read before the* RADIO SECTION 12*th January*, *the* SOUTH MIDLAND RADIO GROUP 28*th February*, *and the* NORTH-EASTERN RADIO AND MEASUREMENTS GROUP 4*th April*, 1955.)

SUMMARY

The paper embraces the principal facets of thermionic valve reliability.

It begins by defining the various types of reliability requirements and then deals with the authors' own work in designing and manufacturing "trustworthy" valves with the objective that they shall give trouble-free service under conditions of vibration and shock. It is stated that, although work was concentrated on only one form of reliability, it has proved to be a major contribution in dealing with other reliability applications. The results so far obtained are discussed.

Emphasis is given to the importance of the contribution of other component makers and of close liaison between valve maker and circuit designer if maximum circuit reliability is to be obtained.

In conclusion, the future trends are outlined.

(1) INTRODUCTION

The electronics industry has grown very rapidly, and with this there has been a large demand for thermionic valves as evidenced by the fact that in Great Britain the valve output has trebled in the past ten years. A considerable part of this has been absorbed by the expanding television market, but there is a significant rise in the industrial usage of valves which is likely to continue because of the many recent advances in the scientific and industrial fields of electronics.

In addition, the need for self-preservation and the continued developments in the Services have shown that electronic devices are a necessity with the modern conception of war. Yet at the start of the Second World War the valve was just a useful adjunct. The vastness of the changes is best illustrated by specific examples. The value of the electronic equipment in the latest-type flotilla leader destroyer approaches 100 times that of its 1940 counterpart, whilst a battleship has a valve complement of 9 000. The electronic equipment in a modern aeroplane costs nearly as much as the complete airframe. The 1940 fighter had 20 valves whilst the latest ones have 600, and a modern bomber needs 2 000 valves to function efficiently.

(2) GENERAL HISTORICAL STATEMENT

The quality standard of the valve designed and manufactured to meet the modest requirements of receivers for home entertainment purposes has been the only one generally available for military and industrial uses, and it is therefore not surprising that the average operational life of such equipments per valve fault has been measured in very small numbers of hours.

Requirements for better-quality valves have arisen from the following sources:

(*a*) *Manufacturers of computing equipment*, since a single pulse at the wrong time can ruin a whole operation because of the presence of a memory which remembers errors as well as the information to be processed. A single installation can use up to 1 800 valves or more, and many investigations have been initiated by the user to improve the equipment reliability by the study of valve failures and the devising of preventative maintenance techniques.

(*b*) *Civil aviation organizations*, which had paramount safety requirements in all-weather operation and in addition found that their economics were seriously affected by the loss of operational time caused by faulty equipment.

(*c*) *Government sources*, where it was realized that the loss of valuable equipment, the success of an operation, and even the turning point of a whole war, could depend on the reliability of a component. Improvements had therefore to be achieved regardless of cost.

Pressure from these sources has resulted in the setting-up of major programmes of development of "reliable" valves in both the United States and Great Britain.

The large-scale effort in this country dates back to 1949, when the Government placed development contracts for the design of "plug-in" replacements for valve types on their Preferred List. Subsequent contracts for the development of new types have called for these to be "reliable" *ab initio*.

(3) CONCEPTION OF THE RELIABILITY PROBLEM

(3.1) Definition of Reliability

The Oxford Dictionary defines "reliability" as "the quality of being reliable," and "reliable" as "that may be relied upon; in which confidence may be put; trustworthy; safe, sure."

The application of this generic word to valves is fraught with many complications, and usually needs a careful association with the end-usage to make it significant.

One attempt at definition has been to state that "reliable" valves are those valves so designed and manufactured as to give continuity of operation superior to ordinary commercial valves.

Another, attributed to Callick, is that "a reliable valve is one having a very high probability that it will operate normally when taken from stock and installed in equipment for which it is intended, and a very low probability that it will fail during subsequent operation in that equipment for some definite period of time."

The valve industry in Great Britain has been disturbed by the fact that the description of certain grades of valves as "reliable" leaves the implication that others not so designated are "unreliable," and thus it has attempted to overcome this by introducing the generic reference "special quality" valves, defined as follows:

A "special quality" valve is a valve which has certain design and manufacturing features making it suitable for use under conditions different from, or in excess of, those experienced in normal radio or television receivers, and when operated under stated or agreed electrical or mechanical conditions it has an acceptable statistically determined expectation of life.

The different electrical and mechanical operating conditions that are possible lead to the following classes of "special quality" valves:

Class (i): Valves to give particularly long lives under conditions where they are not subject to appreciable mechanical shock—e.g. repeaters and computers.

The authors are with Standard Telephones and Cables, Ltd.

Class (ii): Valves to give normal lengths of life under moderate conditions of vibration and shock—e.g. equipment for the Armed Forces, civil aircraft, mobile communications and some industrial uses.

Class (iii): Valves to withstand particularly severe conditions of vibration and shock where comparatively short lives can be tolerated—e.g. projectiles and guided weapons.

Class (iv): Valves to give high electrical stability and normal lives under conditions where they are not subjected to appreciable vibration and shock—e.g. d.c. amplifiers.

Note: These classes must not be regarded as mutually exclusive, and combinations of the attributes inherent in each class may be incorporated in a particular type.

The major emphasis in the paper relates to work done on reliability as defined by Class (ii), but it is fortunate that the improvements so made represent a substantial contribution towards better performance in the other classes.

(3.2) Economic Conception

As explained[100] elsewhere, for a long while valve engineers have considered that the standard of valve reliability achieved bears a close relationship to the price the customer has been willing to pay.

The acceptance of the fact that there were many usages where the low-cost limitation imposed by the commercial radio and television market does not apply has permitted the valve engineer to embark on this quest for the better valve.

The authors have evolved a practical objective in which the standard required is defined as the highest reliability which can be obtained, commensurate with the ability to manufacture by mass-production methods.

(3.3) Definition of Requirements

(3.3.1) General.

Analyses of the failures occurring in service showed that these were greatest when valves were used under conditions of vibration and shock. Some valve manufacturers have reacted to this realization by adding strengthening members to existing valve structures, but it has often happened that the modified valve obtained by this arbitrary means is not necessarily better able to withstand onerous conditions of usage.

The authors have found that the sounder method has been to use a logical and scientific approach which may have taken longer in time but has produced more satisfactory and positive results. Thus a start was made by drawing up specifications which described and correlated the conditions of field usage. Subsequently, laboratory equipment was designed so that a comprehensive study of the mechanical and electrical parameters of typical valves could be made.

(3.3.2) Mechanical.

A mechanical specification could be established only by investigating the vibration and shock conditions which were encountered in equipment. This was done in 1949 by a Government panel which proposed a range of tests to be used as target mechanical requirements for the development of reliable valves. After joint discussions with the valve manufacturers, the final agreed mechanical testing specification for reliable valves was written and has been given in a paper by Hunt.[64]

(3.3.3) Electrical.

The basis of the development programme for valves of increased reliability was that these were to be electrically indistinguishable from the original prototypes and therefore the conditions and limits of the appropriate test specification would apply to the reliable valve. However, as the work proceeded it was realized that in order to take full advantage of the improved mechanical performance in terms of useful life, it was desirable for the electrical parameters to be held to closer tolerances and some reduction of characteristics drift was envisaged. It was also considered that some reduction in electrical noise and microphony would result. Thus the electrical requirements were defined as:

(*a*) Fulfilment of the conditions of the existing test specifications.

(*b*) Reduction of characteristic spread.

(*c*) Minimum drift of characteristics, particularly in early life.

(*d*) Reduced electrical noise and microphony.

(3.4) Life Considerations

Published information[26,49] on the life performance of valves has dealt with many facets of this complicated problem, but in most cases it constitutes factual data resulting from quantity testing with deductions based on an empirical approach.

Reliable detailed data on adequately controlled tests are difficult to obtain, and pending the availability of this it has been thought worth while to attempt a theoretical appreciation so that suitable mathematical formulations shall be available for the effective assessment in due course.

By assuming various failure rates it is possible to evolve failure curves for each by considering the valve failures in a finite population of valves as subject to treatment as a stochastic process. As an example, Fig. 1 illustrates the constant "death rate" case which gives an exponential type of decay curve, and also shows the serious decrease of life expectancy when close tolerances are imposed on a particular characteristic.

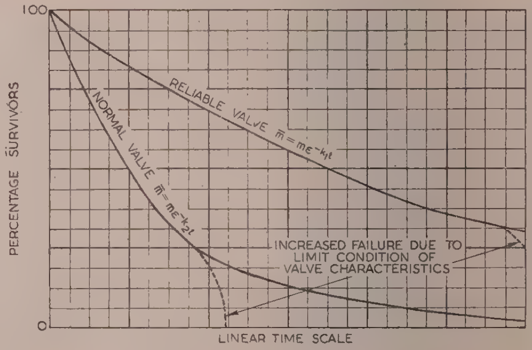


Fig. 1.—Survival curves of valves.

The mathematical treatment is outlined in Appendix 13.1.

(4) EVOLUTION OF MECHANICAL TESTING EQUIPMENT

(4.1) Vibration Testing Equipment

The authors first met the problems of designing valve vibration equipment early in 1949 in connection with the failures of double-triode valves used in an auto-pilot equipment. The rejection of some mechanical defects was found possible by mechanically shaking valves on a vibrating platform mounted on an electric motor, which, in turn, was rubber-mounted on a base plate, the motor having an out-of-balance weight on its spindle. However, with this type of apparatus it was very difficult to deduce the forces applied to the valve, and a more conventional mechanical-vibration table was devised in which the table moved horizontally between guides and was motor-driven from an eccentric cam. With this machine the amplitude of vibration is fixed by the dimensions of the cam, but the frequency and

acceleration parameters are independent. Thus it was not easy to decide the relative stressing effects of acceleration and frequency.

Experiments were then carried out on a third form of mechanical test equipment, often described as a "bump testing" machine, which was essentially a hinged board, oscillated by a cam drive. Because it subjected valves to many repetitive shocks it revealed mechanical weaknesses, but it suffered from the disadvantages of the first equipment described because it did not allow precise investigation. Such mechanical testing machines can be of value, particularly in the case of small-quantity manufacture, but the information derived from them cannot easily be evaluated for design purposes since they serve as a go/no-go test rather than as a controlled deterioration test. Wear on the various moving parts causes distortion of the waveshape with consequent misleading results and also makes difficult the correlation of equipments built by different manufacturers.

About this time the detailed Government mechanical specification was evolved and it was realized that, in addition to the limitations mentioned above, it was unlikely that mechanical equipment could be produced to give reliable operation at the desired higher frequencies, and therefore design work was concentrated on using some form of electromagnetic transducer.

After many experimental models were made, the final equipment consisted of an automatic recorder vibration unit capable of fatigue-vibration testing more than 100 miniature type valves at a time. The table was excited from a De Havilland Type 1/D3 moving-coil vibrator driven by a 100-watt amplifier and oscillator. Monitoring was achieved by bringing the anodes of the valves out separately, amplifying the noise voltage developed across an anode load resistor and referring this to the preset grid voltage on a gas-filled rectifier, which operated relays to indicate "good" or "bad" valves on a roll of special paper. Such equipment had to be free from vibration-table resonances, and it was found necessary to replace solid-steel tables by others using a composite construction of aluminium.

Recent work on valve design has indicated that unless higher accelerations are used for fatigue testing, the failures will be so few that the test will not be sensitive to changes in quality. It has been found that testing systems employing a resonant bar will give up to $50g$ with a load up to 8 lb, and a prototype vibrator is being used to establish whether these higher accelerations increase the severity of the test without producing other catastrophic failures not found in normal valve-life under arduous conditions.

(4.2) Shock Testing

Advantage was taken of the fact that the United States had standardized an equipment known as the "Taft–Pierce fly-weight high-impact shock machine." To make it more suitable for the testing of radio valves a reduced scale model was constructed. A photograph of this machine has been published.[65]

(4.3) Centrifuge Testing

For accelerations above $1\,000g$, centrifuge testing equipment is necessary, but it is not considered that such apparatus gives better information than is obtained by vibration methods.

(4.4) Resonance Search Testing

A simple form of electromagnetic transducer using a variable-frequency oscillator was used initially but was superseded by a more elegant apparatus described elsewhere.[65]

With this equipment the valve under test can be vibrated in any direction, but it is usually mounted parallel to the minor axis of the grid structure and is operated in class A conditions with a suitable anode load resistor. The a.c. noise voltage at the anode of the valve is displayed on a cathode-ray tube, and by traversing the frequency spectrum it is possible to make a film record of the peak noise output against frequency.

The equipment is limited in its upper frequency by the efficiency of the transducer, and for investigations above 3 kc/s other forms of excitation have been sought. One possible method[100] which has proved very satisfactory consists in mounting the valve in front of a loudspeaker energized by a variable-frequency oscillator and selecting the resonant peaks in the test-valve output as the frequency is raised to the cut-off point of the loudspeaker. By this means a complete resonance spectrum up to 15 kc/s can be produced for any valve.

(5) STUDY OF EXISTING COMMERCIAL VALVES

(5.1) General

Over the years there has been a steady increase in the number of valves used per equipment, and the imperative need to get more and varied equipment into the smallest possible space led to the desire for smaller valves. Comparatively small valves had been made for specialized uses more than 20 years ago, and there was some small commercial production of hearing-aid valves in the late 1930's, but the real impetus on size reduction came in 1941, when valves were required for use in proximity fuses. The valve maker saw in miniaturization a basic way towards lower costs by savings in materials, but the much more important advantage of increased ability to withstand vibration was not appreciated until later. The ratio of strength to mass increases with reduction of size, whilst the low centre of gravity of the mount and the short mica-to-mica distance favours ruggedness against shock. Furthermore, the smaller mass and size make the glass structure inherently less prone to breakage on impact and glass strain. Thus it was logical that work should be concentrated on small all-glass valves.

(5.2) Results of Mechanical Testing

As an integral part of the early development of reliable valves, normal commercial valves were subjected to mechanical testing and the rate and type of failure were observed.

The commercial equivalent of a trustworthy type was used as a datum line (or control) in experiments on reliability, and the success of the design was measured by reference to the failure rate of the commercial equivalent.

Since no previous study had been made on vacuum tubes under these conditions of vibration and shock, little was known of how normal valves would withstand the proposed tests. All that had been established was that valves were unsatisfactory in the field under conditions which were reputedly simulated by these tests.

Shock testing showed two effects. First, the initial shock tended to cause a general reduction of anode current, and subsequent shocks had no effect, this phenomenon being most noticeable in valves with high mutual conductance. Secondly, the electrode distortion caused by shock testing was most severe when the direction of shock was along the length of the valve or along the minor axis of the grids. With commercial valves, shocks greater than $1\,000g$ caused complete electrode collapse.

Centrifuge testing revealed that there was a critical acceleration below which long periods of testing would have no effect, but above which complete failure would result in a few seconds. The limiting acceleration was lowest for valve designs where the anchoring of the assembly to the glass base was weak.

The most severe form of test was found to be that of vibration fatigue where the types of failure analysed included all those observed on the other tests. Initially, and in the first few thousand vibrations, there were a number of short-circuits and dis-

connections, whilst gas and low-emission failures due to mica dust appeared later. Early tests were done under class A conditions, but it was established as equally effective, and simpler, to omit the h.t. supply and to operate with heater supply only. The rate of failure varied from type to type, and such failure curves all had the same general shape although the tests were carried out at different frequencies. When replotted with the numbers of vibrations as the abscissae (Fig. 2) such curves then

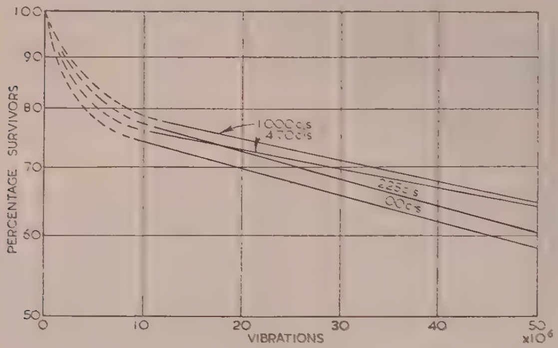


Fig. 2.—Survival rate as a function of the number of vibrations.

became almost coincidental. Thus the total number of applications of the force seem to be more important than the actual vibration frequency used, and this evidence is disturbed only if any resonant frequency of the valve system coincides with the frequency of vibration.

Some commercial valves were found to have been very well designed mechanically. For instance, valve type 6AL5 survived much better than the 8D3/6AM6; of a number of 6AL5 valves only 20% had failed after 25 000 vibrations, whereas a similar number of 8D3 valves had all failed at 15 000 vibrations.

(5.3) Correlation with Field Failures

It was not easy to get correlation of these vibration-life results with those available from the field. In the early stages of the development work, no controlled experiments were being carried out and the only information obtainable showed the type of failure without any life indication.

Table 1 gives the frequency of occurrence of particular faults, the majority of which were short-circuits and disconnections similar to those occurring early in the vibration period.

Table 1.—Distribution of Field Failures

Fault	Computer (stationary use)	Services (mobile use)
	%	%
Electrical	68	61
Mechanical	25	16
Glass	7	23

(5.4) The Application of Selective Tests to Commercial Valves

By employing special mechanical tests in addition to the electrical characteristic tests required to maintain a strict control of the product, it was found possible to select from normal commercial production a proportion of valves that would give a better performance than the unselected product. This was not at variance with the essential idea of valve reliability but rather supported it.

When used in equipments not subject to mechanical vibration or shock, valves selected in this way give a high standard of performance which may be termed reliable, whilst under vibrational conditions such valves show a considerable reduction of catastrophic failures. This point has been proved both by field results and from special vibration tests conducted on the valves themselves.

Experience with the selection of commercial radio valves by special tests has shown that it has formed a satisfactory interim measure before the fully reliable valves become freely available.

In addition, there are many of the older valve types required for maintenance purposes only, and because it would be impractical in time and expense to redesign these for full reliability, a programme such as this represents the best compromise.

(6) LABORATORY WORK TO ACHIEVE RELIABILITY

The laboratory work to achieve reliability can be classified in three sections, namely (*a*) mechanical aspects of valve design, (*b*) electrical considerations, and (*c*) glass technology.

(6.1) Mechanical

The results of mechanical tests on commercial valves and information received from analyses of field returns have revealed the mechanical shortcomings of the structure and components of existing valves.

(6.1.1) Valve Structural Considerations.

Fatigue life-tests on the harmonic vibrating table showed that the most serious cause of valve failure was the evolution of gas, resulting in cathode poisoning or breakdown. The cause of this gas evolution was traced to frictional movement between the mica insulators and the valve envelope and components, and the elimination of this has been the most important contribution to valve longevity under conditions of vibration and shock.

To prevent longitudinal movement of the cage, careful attention must be paid to the design of the cage-stem connections. The aim is to achieve as many direct connections as possible between the cage and the stem and to locate them close to the periphery of the cage in an approximately equilateral arrangement. All stem wires should be kept as straight as possible to minimize "springing" under longitudinal forces. Anode lugs, shield lugs, heater bars and mica clips can be used to achieve the rigid cage-stem conditions. All such anode and shield lugs are either bent over mechanically to hold the cage firmly to the bottom mica or they are taped. Similar conditions apply to the top mica, and the result is a mount of extreme rigidity which prevents relative axial motion between mount and envelope.

The problem of lateral movement inside the bulb itself has to be taken care of by the provision of bulbs made to close internal tolerances and an accurate control of mica-bulb interference by the judicious dimensioning of the snubbering micas.

The general trend in structural design has been towards shorter structures to lower the centre of gravity of the valve and to raise its inherent resonant frequencies. Shorter components will have a lower ratio of length to moment of inertia, which will give a smaller amplitude of movement for a given energy input with a resultant reduction in resonant noise output from the valve.

Many vibration-life failures are caused by the getter. Loose getters can give rise to short-circuits and random noise, and faulty directioning of the getter spray can cause noise through electrical leakage across the mica. Also, cantilever mounting of normal "horseshoe" getters gives rise to a low-frequency noise of about 1 kc/s. Not only have improvements to welding techniques been made, but getters are now welded in two places where

possible, and these welds are widely spaced to achieve beam mounting, thereby inhibiting vibration and reducing the possibility of looseness. Where mount height permits, an additional "splash" insulator is incorporated to prevent getter spray causing leakage between electrodes on the main insulator, and in the event of weld failure this prevents the loose getter member from causing a short-circuit between electrodes.

Two important considerations which guide structural design are that the design of insulators, anodes and shields must be such as to promote the rapid exhaust of gas, and as ease of assembly is a direct contributor to reliability in the final valve, attention must be given in the design stage to the possibility of jig assembly of the components into sub-assemblies, and thence into the final mount. Fig. 3 illustrates some of the principles described.

Its use is sometimes precluded because of electrical characteristic requirements, and it is then necessary to use other sections in the following order of preference: oval flattened on the minor axis, flat major axis produced from tubing, rectangular, oval.

Considerable research has taken place to find suitable nickel-alloy cores with improved hot strengths without sacrificing other necessities such as satisfactory emission current-densities obtainable with reasonable treatments.

Another cathode requirement which has been in the forefront in recent years is that there shall be a low rate of interface growth. In Great Britain the normal core material has had magnesium as the main impurity and therefore this country has not been so much embarrassed with this phenomenon as others which use silicon additive.

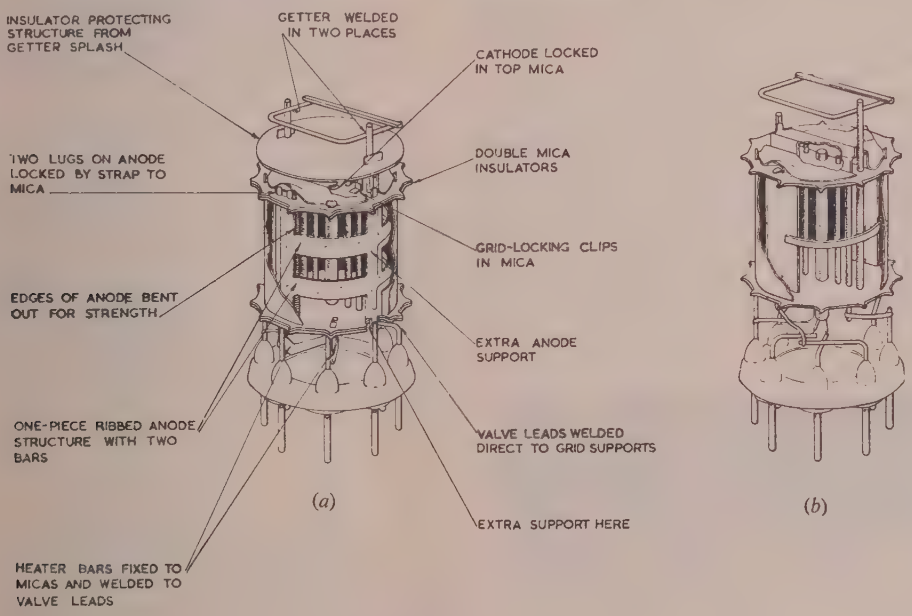


Fig. 3.—Typical valve construction.

(*a*) Trustworthy. (*b*) Ordinary.

(6.1.2) **Heaters.**

To achieve maximum reliability it has been necessary to use tungsten exclusively and to design for the maximum volume of heater core material so that the heater temperature can be kept as low as possible consistent with operating requirements.

In all circular cathode designs the double helical heater has been used, whilst with valves having shaped cathode sleeves the hamp-coil heater has been chosen as the next best design for reliability.

Special coatings made from mixed aluminas have been evolved to ensure that the coated heater will not fracture during life and will have adequate heater-cathode insulation.

(6.1.3) **Cathodes.**

The rigidity of the cathode itself is very important if maximum valve reliability is to be achieved.

A study of the mechanical strengths of various sleeve cross-sections has shown that the circular one is the strongest because it has the greatest moment of inertia of cross-section for a given periphery. It is also preferred because it is the easiest to manufacture and diameter tolerances as small as $\pm 0 \cdot 0003$ in can be maintained.

It is fortunate that the alloy metals contributing to increased strength have also given lower rates of interface growth. Comparative values obtained after 500 hours in the cut-off condition at slightly elevated cathode temperatures are:

Material	SiNi	"O"Ni	Experimental alloy
Interface resistance, ohms/cm^2	80–2 000	20–80	1–5

Tests up to 2 000 hours show that the experimental alloy does not give any appreciable increase in interface resistivity after 500 hours, but with "O" nickel, this resistivity rises to 800 ohms.

(6.1.4) **Other Components.**

The grids used have been improved versions of normal commercial grids, but oval profiles are preferred as being the strongest and are used with minimum-length laterals and wires of the largest possible diameter to achieve the highest resonant frequencies and the lowest amplitudes of resonant vibration.

Anodes and shields are designed to obtain the maximum insulator-component rigidity, and increased mechanical strength is achieved by the use of the strongest profiles and by the scientific positioning of ribs.

It is unfortunate that mica is still the most suitable material for insulators, and there is a need for an alternative material which does not split or flake or suffer from serious frictional wear.

(6.1.5) Insulator-Component Locking.

Adequate attention to the locking together of the various components is essential if mechanical rigidity and subsequent freedom from vibration are to be obtained. Such locking has to take into account the fact that whilst restricted movement is necessary during operation there must be sufficient freedom for expansion under heating.

Design techniques for securing cathodes, grids and pressed parts to meet these requirements have been evolved.

(6.2) Electrical

(6.2.1) General.

The limiting sensitivity of most electronic equipment is the signal/noise ratio which, it is accepted, can be improved by suitable choice of circuit and valve. However, the authors have found that the noise output of valves is predominantly caused by mechanical loosenesses and that noise due to electron movement is a second-order effect. Under adverse operational conditions the noise output may swamp low signals or may provide false information at the output, and therefore an important feature of reliable valve design is to minimize the noise output under vibration.

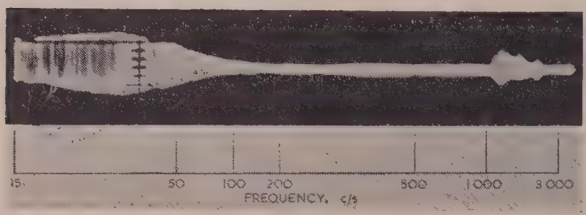


Fig. 4.—Resonant noise output from a valve.

Fig. 4 is a photograph showing the resonant noise output from a valve as obtained on the resonance test set. Analyses of many such results show that the noise output spectrum can be divided into two distinct regions:

(i) A low-frequency noise which is fairly independent of frequency up to about 500c/s and varies in amplitude from valve to valve.

(ii) Discrete bands of high Q-factor resonant vibrations at frequencies generally above 800c/s.

(6.2.2) Low-Frequency Mechanical Noise.

The resonance search equipment previously described is overelaborate for the investigation of low-frequency noise, and a simpler and more flexible system has been produced using a proprietary vibrator driven by a 100-watt amplifier. This arrangement allows small masses such as miniature valves to be vibrated at accelerations up to 30g. The valve is operated in a class A circuit, and the noise output is developed across the anode resistor and presented across a 15-ohm load.

Fig. 5 shows the distribution of noise output in a batch of commercial valves; the long tail of valves with high noise values is typical of normal domestic valve production.

The noise output spectrum at low frequencies is complex and

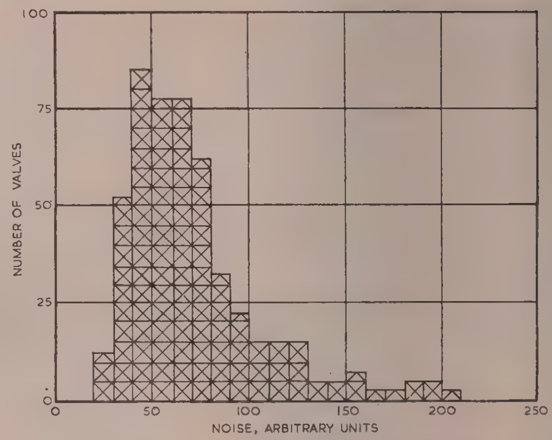


Fig. 5.—Low-frequency noise distribution.

has been investigated by vibrating valves deliberately made with certain loose components. It has been found that vibration in the vertical plane produces a spectrum which has little relationship to the loose component and is distorted by spurious effects such as electromagnetic deflection of the space current by the vibration field. Vibration across the major axis is much more definitive and is therefore invariably specified for test purposes.

It has been shown elsewhere[102] that for a given displacement the cathode produces a greater change of anode current than any other electrode, and thus the satisfactory fixing of the cathode is of the highest importance. In Fig. 6 is shown a typical frequency

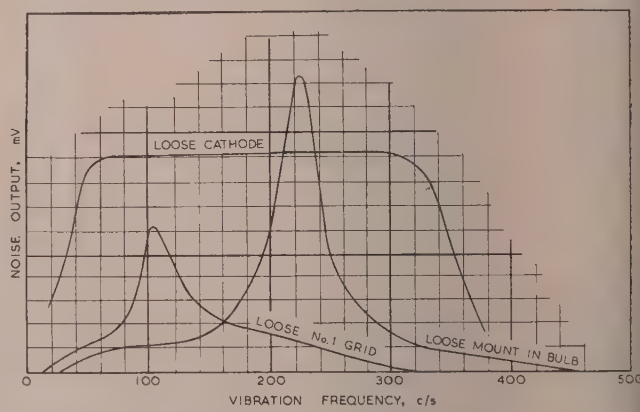


Fig. 6.—Noise output due to loose components.

spectrum which is due to cathode rattle and is flat-topped because of saturation effects. The same diagram illustrates the noise from a No. 1 grid and also the effect of mount resonance.

With better valve design and manufacturing methods, the level of low-frequency noise at the standard measuring acceleration of 2–2½g becomes very low, particularly in sub-miniature valves, and to investigate the possibilities of further improvements it has become necessary to use higher values of acceleration. Tests are now being made at 12–15g but there is a lack of correlation between noise levels at 2 and at 12g. There appear to be three categories as indicated in Fig. 7, namely:

(*a*) Valves which show low noise at 2g and 12g because the electrodes, particularly the cathode, are locked effectively and remain so for many hundreds of hours of vibration.

(*b*) Valves which show low noise at $2g$ but are noisy at $12g$. Here the electrode locking is effective initially but fatigues under vibration.

(*c*) Valves which are noisy both at $2g$ and $12g$ owing to unsatisfactory locking.

Valves of category (*b*) cannot be guaranteed for continued use even at $2g$, and there is a case for testing all valves at, say, $12g$, to afford adequate cover for lower-g working.

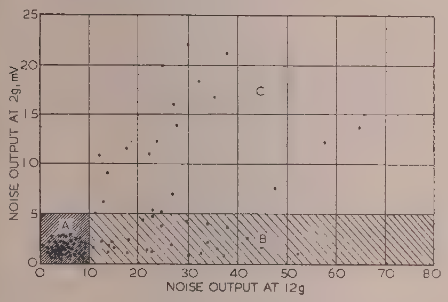


Fig. 7.—Comparison of valve noise at $2g$ and $12g$.

(6.2.3) **H.F. Resonances.**

At frequencies above 800c/s the noise output from a valve is essentially sinusoidal because it is due to inherent resonance of the valve components.

Most of the resonances are due to the grid supports and lateral wires, the latter giving discrete bands of resonant frequencies due to the slight variations in the dimensions of the individual lateral wires of a grid (see Fig. 8). Consideration of the modes of vibration of the various components has made it possible to evolve formulae for determining the resonant frequency of each element. Thereby the engineer has been able to pursue the theoretical design of valves which have to be free of resonances in certain frequency bands.[102]

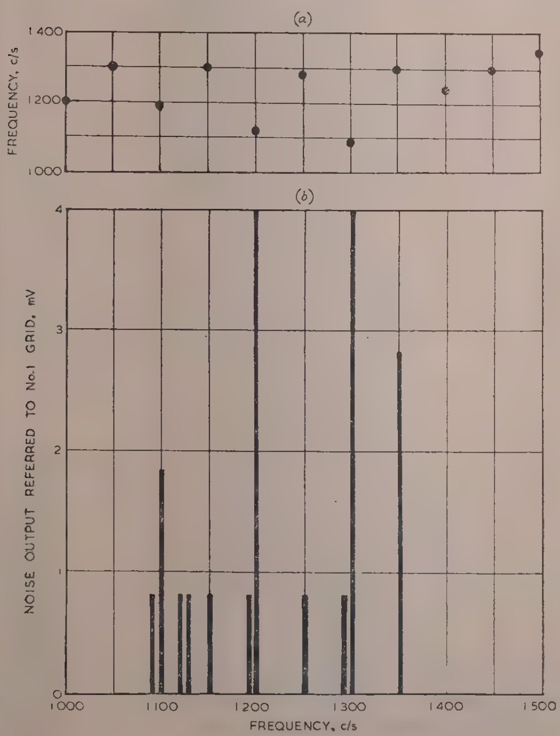


Fig. 8.—Resonances of individual grid lateral wires.
(*a*) Resonant frequencies of individual laterals along one side of a grid.
(*b*) Noise spectrum due to the grid lateral wires.

(6.2.4) **Microphony.**

A quantitative assessment of valve microphony may be made by employing a loudspeaker system with a suitable response characteristic and using acoustic feedback principles. The valve under test is mounted in front of the loudspeaker, and its output is fed back into a calibrated amplifier which feeds the loudspeaker. By adjusting the position of the valve and the amplifier gain, the circuit can be made to "ring" at the preferred frequency of the valve. Measurement of the amplifier gain to achieve the threshold of acoustic feedback together with the value of the preferred frequency gives a measure of the "goodness" of the valve.

This procedure is used to set the initial design level and can be employed subsequently as a factory test; in practice, it gives consistent results when operated by semi-skilled personnel.

(6.2.5) **Valve Ratings.**

It is well known that improved reliability is achieved by using valves below their maximum ratings, but it is important that all specified ratings shall be correlated with the ambient conditions under which the valve is likely to have to operate. Ambient conditions of high temperature and reduced pressure cause the bulb and electrode temperatures to rise, with the following possible effects:

(*a*) A decreased life expectancy due to the loss of cathode emission and/or heater failures.
(*b*) Grid emission due to contamination, excited at the higher electrode temperatures.
(*c*) Gaseous condition in the valve due to gases released from overheated electrodes and the valve envelope, and by the reversal of the getter action.
(*d*) Electrolytic conduction of the glass between pins, especially in all-glass rectifiers.
(*e*) A drift of valve characteristics.
(*f*) An acceleration of interface growth.

To cover the full range of ambient conditions, rating charts must be prepared for each valve type.

(6.2.6) **Stability of Valve Characteristics.**

The problem of the stability of valve characteristics during operation is of great importance to designers of most electronic equipments and in particular in the design of d.c. amplifiers. There are two problems, the long-term life performance of valves and the short-term effect where the electrical characteristics drift over fairly short intervals of time, i.e. hour by hour. This drift is due to mechanical and electrical causes.

The improved mechanical rigidity of reliable valves has resulted in a marked increase in stability of electrical characteristics because of the accurate maintenance of inter-electrode spacings. This forward step has been further assisted by attention to cathode power ratings and heater designs and has made possible double triodes of the 6158 class which show improvements as illustrated by Fig. 9.

(6.3) **Glass Technology**

(6.3.1) **General.**

Valve making involves extensive glass-working operations which have evolved from lamp-working practice and are the results of long-established practical experience. There is little published literature dealing with the technological approach whereby theory is applied to practice, but this has been dealt with elsewhere by one of the authors.[65]

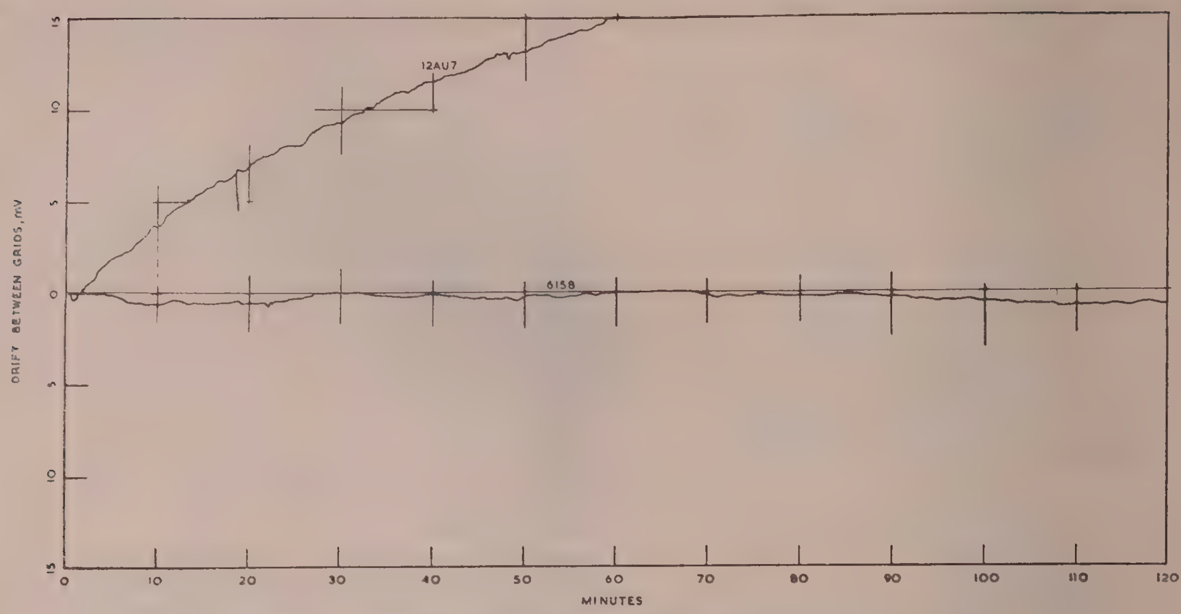


Fig. 9.—Typical 2-hour stability curves for valves of types 6158 and 12AU7.

A scientific attack has been the study of the stress distributions in the bases, the examination of the type of base cracks in manufacture, and the evolution of factory tests which can simulate the glass faults found in the field.

The results of these studies are utilized in a control method based on thermal shock testing which has been established for the sealing-in and pumping processes.

(6.3.2) **Thermal Shock Tests.**

The theory of the thermal shock test is that if the stresses localized at a point are likely to be dangerous to life (because of delayed fracture effect), the thermal shock will set up temporary stresses which, added to the localized stresses, will cause immediate failure from that point. Thus, the thermal shock test "searches" for dangerous degrees of tension, the sudden heating searching the inside surface, and the sudden cooling searching the outside surface.

Two of the thermal shock tests applied are already familiar in the valve industry. The first, called the "A" test, consists of plunging a cold valve into boiling water and holding it there for ten seconds. In the second, called the "B" test, a tapered metal plug is inserted between the valve pins and, with its pins slightly forced apart in this way, the cold valve is plunged into boiling water and held as before. After being immersed for ten seconds the valve is withdrawn and allowed to cool freely to room temperature with the plug still fitted between the pins.

The "A" test is a simple detector of internal tensions, whilst the "B" test is more complex, reacting not only to internal tensions, but also to such other factors as pin stiffness, and to some extent, external tensions.

A comparison between the frequency of occurrence of various kinds of cracks found in the field and those obtained on the "A" and "B" tests showed little correlation. Work has therefore been continued to find a test which would give this very important information. Thus has evolved the "T" test, whereby the valves are slowly raised to the temperature of boiling water and are then suddenly plunged into cold water. Such a test gives results closely agreeing with field failures.

(7) MANUFACTURE

(7.1) General

The authors have shown how the engineer can design reliability into a valve by calculation and by special testing carried out on laboratory equipment. However, an average valve has 7 glass-metal seals and 33 welds with over 800 separate and distinct manufacturing steps to convert the raw material into the finished product, and the production engineer has the difficult task of manufacturing mass production quantities of such complex articles with the minimum variation of mechanical, chemical and human tolerances associated with the materials and the processes.

The problems of reliability resolve themselves into greater efforts to control the materials, the processes and the operator variability.

Inferior materials may lead to failures in meeting the stringent mechanical tests, in which case the whole of the production is suspect. Lack of control of processes will lead to extreme variations within the manufacture and result in a proportion of poor-quality valves, whilst operator variability can produce valves which are potential failures for mechanical faults.

(7.2) Valve Assembly

There are two schools of thought regarding the place in which special-quality valves should be manufactured. One advocates an entirely separate location from the commercial types, but much may be said in favour of their assembly in the centre of the main domestic-valve groups, so that, with strong supervisory control, the effect of the lessons learned will have a large psychological influence on the whole factory. This point is doubly important when it is realized that in the event of an international crisis very large numbers of special-quality valves would be demanded.

The production of quantities of valves which must show extreme uniformity of performance throughout a severe mechanical and electrical test sequence entails a rigid control of assembly to ensure that manufacturing variations are minimized and operator errors are reduced to negligible proportions.

The establishment of such high quality demands continuity of

production over long periods, and the corollary to this is that the valve type diversity shall be limited as much as possible. Short-term runs will inherently yield poorer reliability. That these ideas of quantity production are correct is illustrated by the shrinkage control chart shown in Fig. 10, in which it will be seen that each time a type has to be reintroduced it starts at a considerably higher shrinkage than the target.

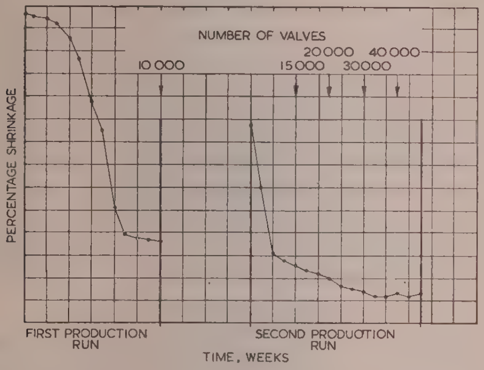


Fig. 10.—Shrinkage graph for 6065 type valve.

One of the most insidious of valve faults is intermittent noise due to lint which arises from foreign matter caught up in the piece-parts. This can come from many sources—e.g. from the atmosphere, from operator's clothing—and as it carbonizes during valve processing it forms variable leakage paths between electrodes. Special processing of the finished valve can reduce lint noise but cannot ensure a permanent cure. Therefore precautions must be taken to exclude lint at every stage of manufacture. Air conditioning of the factory itself will reduce to a minimum the material which can enter from external sources, whilst covered benches for the critical operations, together with the storage of components in specially designed trays with glass covers, give additional safeguards against contamination. All operators are provided with nylon overalls to preclude clothing fluff getting into the valve structures.

The actual assembly operation is the most critical one, and special thought has been given to assembly-bench design. The bench tops used are covered with a smooth shiny plastic material and are free from obstructions other than welder heads. The bench itself is of enamelled metal having a cover with a glass front under which the hands of the operators work. The assembly chamber is pressurized so that there is a slow flow of air outwards, thus preventing ingress of lint.

It is important that operator fatigue be studied. The height of the bench is chosen to be the same as that of normal home furniture, and the operator sits comfortably in a chair in a position where movement and strain are minimized. Illumination is by lighting concealed under the cover to obviate shadows and eye-strain.

The most difficult variable on assembly operations is that of resistance welding. A high proportion of potentially faulty valves can result from quite a low proportion of faulty welds, and to achieve the greatest security from welding troubles a completely new welding head with controlled timing equipment has been devised.[65]

Initially, the assembly of trustworthy valves was carried out on a time-work basis with no incentive towards speed. However, it was found that this was so alien to the mass-production outlook obtaining in valve manufacturing that a change was made to operator teams with the assembly sequence broken down so that the skill required was reduced as much as possible. Frequent changing of welder setting was avoided by the provision of double welder positions. This approach permitted a quality-control system to work on each assembly position, and it has proved to be a more stringent control than 100% final inspection. It has been possible to introduce an incentive scheme based on quality and quantity, and a study of the results has revealed that, when an operator is given a simple sequence of jig-aided operations, work begins to flow at her natural rate with maximum efficiency.

The evolution of special aids for the semi-automatic assembly of the cages has contributed much to reliability. It has been possible to assemble a pentode with a cathode–grid clearance of 0·005 in (0·12 mm) with a new operator of two weeks' training at a speed greater than the most experienced operator using the old methods. The improved quality obtained was so apparent that it was possible to mix mounts and then to divide them into two distinct groups in the mechanical test routine.

(7.3) Testing

The design work on trustworthy valves has been concentrated upon the stringent mechanical requirements, and it has been shown how special test-machinery has been developed for checking under these conditions. With adequate statistical control of materials, piece-parts and processes, including valve assembly, satisfactory valves can be made at an efficient rate, but the achievement of failure rates as low as 2% per 1 000 hours is not dependent solely upon structural design and the control of the manufacturing unit. Good design and manufacturing controls combine to ensure that the manufacturing variations will be small and that there will be few random faults or errors, but they cannot ensure their complete elimination. It is imperative, therefore, that the form of valve testing adopted shall take into account both "manufacturing variations" and "manufacturing errors." The first of these can be dealt with by the well-established methods of quality control, but the second group is more difficult to check since they are basically a function of the efficiency of the whole valve-manufacturing plant itself and are independent of the particular valve type that is being tested. It is important that this point should be understood before any form of final testing is devised, since without this appreciation it is possible to evolve a series of unwieldy tests which can make large-scale production impracticable. The development of such testing procedures has been described by Rowe and Welch.[100]

(8) THE CUSTOMERS' CONTRIBUTION TO OVERALL EQUIPMENT RELIABILITY

(8.1) General

The valve manufacturer has faced up squarely to the problems involved in making a reliable valve, but it must be appreciated that valves have not a high safety-factor and often fail through the fault of other components.

The reliability of an equipment consisting of a number of separate components is the product of the individual reliabilities of these components. Fig. 11 demonstrates the reliability required from the other components as a function of valve reliability.

Even the normal television receiver has about 500 electrical components of which only 20 are valves, and it should be realized that, whilst valves have received much attention because of their large contribution to failures, a major improvement in these will reveal many weaknesses in other components.

(8.2) Equipment Designer

The equipment designer can do much to improve the reliability of his apparatus by using valves correctly. The rate of failures of

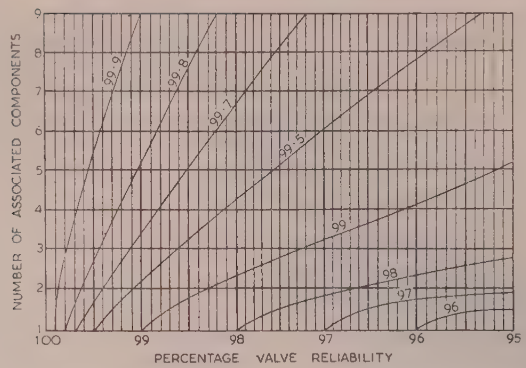


Fig. 11.—Comparison of reliability of equipment components.

specific valves in different equipments can vary by a factor of 10, and this can best be minimized by co-operative effort between the designers and the valve manufacturers. It is accepted that the versatility of the valve itself gives scope for the circuit designer's ingenuity, and that there will always be new methods devised to meet new demands. All that the valve manufacturer asks is that full advantage shall be taken of his intimate knowledge of the idiosyncrasies of valves. Valves are defined by specifications, but these can cover only the applications known and visualized at the time the valve is introduced. Close collaboration can ensure that all valves which meet the test specification will perform satisfactorily in service and will enable the valve manufacturer to make adequate checks to cover any usage of special characteristics.

By this means, the best compromise is found whereby the most suitable valve is used from the standpoint of published and unpublished characteristics and continued availability, and that the best-known circuit techniques are utilized to accomplish that purpose.

(8.3) Equipment Manufacturer

It may not be realized how much reliability can depend on matters more mundane than circuit technique and valve characteristics. The valve is a glass article and should be treated as such—glass is severely weakened by the minutest of scratches—and jumbling valves together in a box, for example, will produce scratching by the nickel pins.

Modern valves, such as miniatures, have a complex multiple glass-metal seal, and leaks result from strains caused by mechanical incompatibility with the valveholders. It is therefore important that wiring jigs shall be inserted into all holders before chassis wiring takes place, and as the valve pins are easily distorted on handling, all valves should be pin-straightened in a proper pin-straightening jig, and not with pliers, immediately before insertion into holders.

In circuit testing, the valve should not be tapped harder than is necessary to check for noise, and the tendency to use a screwdriver for this purpose is unfortunate.

It may be thought that some of these comments are irrelevant, but experience has shown that such practices contribute materially to setting up conditions causing delayed fractures which occur some time after the installation of the equipment.

The valve manufacturer's outlook on the correct usage of valves is summarized in British Standard Code of Practice C.P. 1005.

(8.4) Equipment User

The actual operator of the equipment should pay special attention to any instructions given, particularly since in many cases it is possible to devise maintenance techniques to achieve the optimum service schedule. When the cost parameters are known, these can be derived by probability calculations.

(9) RESULTS OBTAINED FROM "TRUSTWORTHY" VALVES

(9.1) General

The work done in connection with reliable valve development and manufacture has been explained. Interest should now be focused on the achievements realized to date, and it is proposed to examine how far the trustworthy valve has met the demands made on it.

(9.2) Production

Approved testing routines have been used as a control throughout development and manufacture. The design approval procedure has consisted in the complete testing of 200 valves of each type, and Table 2 shows the results obtained. It will be

Table 2.—Results from Application of Approved Mechanical Tests on Batches of 200 Valves of each type

Test	Valve types		
	VX.7075 6065	VX.7083 6064	VX.7076 6058
Initial resonance search test: 15–500 c/s at 2g	2% of the valves showed slight low-frequency noise	½% of the valves showed slight low-frequency noise	No resonance points of low-frequency noise
Fatigue test: 30 hours each in 3 positions of mounting at:— 25 c/s, $\pm 0\cdot 04$ in 50 c/s, $\pm 0\cdot 01$ in 170 c/s, $\pm 0\cdot 002$ in	Two rejects. One cathode-tail disconnected. One valve cracked tip-off	One reject: getter disconnected	No rejects
Shock test: Total of 20 blows; 5 blows in 4 positions at 500g	Two rejects. One cathode-tail disconnected. One grid 2 to grid 3 short-circuit	One reject: getter disconnected	No rejects
Final resonance search test: 15–500 c/s at 2g	A slight increase of low-frequency noise of valves exhibiting noise previously	A very slight increase of low-frequency noise	No low-frequency noise
Summary of results	4 failures: 98% satisfactory	2 failures: 99% satisfactory	No failures: 100% satisfactory

observed that the experience gained on the early types has been applied with success on subsequent developments.

In production, apart from an initial period when operators and engineers alike were new to the project, the shrinkage has been held at a percentage lower than is usual in the industry. A graph shows the progress made with one valve type (see Fig. 12).

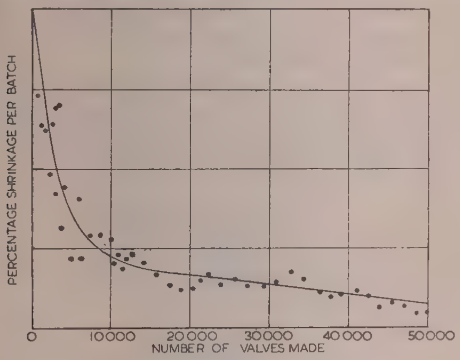


Fig. 12.—Shrinkage chart for type 6064 valve.

The improvement in the techniques of manufacture has led to a narrowing of the spread of the electrical characteristics, and the quality-control charts which are kept for all electrical parameters indicate that the variation in each batch of reliable valves is smaller than that of the corresponding commercial valves.

Naturally, this improvement in quality contributes to an improved stability on static life test. Histograms have been drawn of the change in mutual conductance over a 500 hour life-test for types 8D3 and 6064 (see Fig. 13). Again, it can be seen that the reliable version has greater stability and maintains its characteristic better during life.

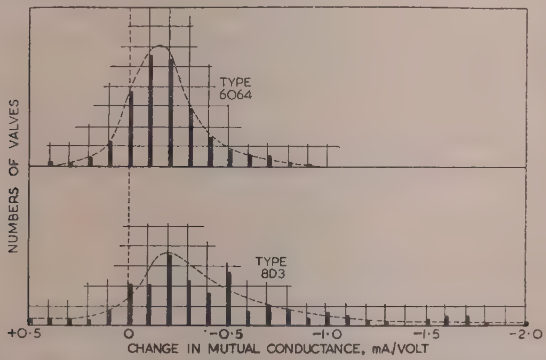


Fig. 13.—Histograms comparing mutual conductance change in 500 hours' static life test for type 6064 and type 8D3 valves.
(*a*) Type 6064.
(*b*) Type 8D3.

However, the reliable valve gives the equipment designer other advantages in addition to a reduction in the spread of characteristics. Particular attention has been paid to the low-frequency noise occasioned by the use of valves, and histograms have been prepared showing the reduction of such noise by the use of trustworthy valves. Fig. 14 shows histograms of results obtained on the 6067 contrasted with those obtained for the 12AU7, its commercial equivalent. It will be observed that the whole distribution has moved nearer the origin, indicating that the improvement is of a fundamental nature.

Glass testing has ensured that the strain patterns induced in

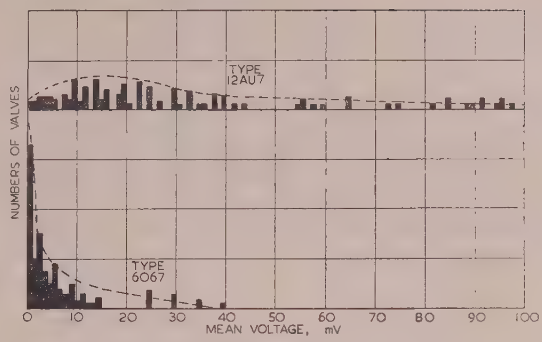


Fig. 14.—Low-frequency noise distribution for type 12AU7 and 6067 valves.
(*a*) 12AU7.
(*b*) 6067.

both base and bulb are such as would minimize the possibility of failure due to a glass defect. Thermal shock testing has resulted in less than 1% failures over all batches for a long period of time, showing that the process control is adequate and continuous.

The confidence one may have in the production processes is such that the rare failures that do occur serve only to focus attention on the normal high quality achieved.

Throughout the production run, considerable numbers of samples have been placed on static life test for the purpose of judging valve performance.

A method widely used both in England and in the United States for assessing valve life is based on a 500 hour life-test. On the completion of this period the average life of the group of valves on test is calculated and expressed as a percentage of the 500 hours as follows:

$$\text{Average life percentage at } x \text{ hours} = \frac{(\text{Life hours for each valve}) \times 100}{x \text{ hours} \times (\text{number of valves started})}$$

The minimum acceptable life performance as given in American specifications is 80% for normal commercial types and 95% for the reliable types.

This conception of life applied to the trustworthy valves under discussion in this paper has provided the following figures:

	%
6065 (CV.131 reliable)	98·30
6064 (CV.138 reliable)	98·42
6058 (CV.140 reliable)	99·21

The figures are taken over the whole production run and include the starting-up period in which there were some early life failures. Corresponding figures taken after production was established show a considerable improvement:

	%
6065 (CV.131 reliable)	99·82
6064 (CV.138 reliable)	99·0
6058 (CV.140 reliable)	100·0

(9.3) Field Results

The results discussed above relate to tests on machines designed for factory and laboratory usage and are meaningless unless there is close correlation between these and field conditions. It is appreciated that with equipment distributed world-wide, often in the hands of semi-trained personnel, it is not easy to organize adequate information. The problem is made more difficult because the substitution of the improved valve removes the cause of the complaint and makes it unnecessary for the customer to incur the expense involved to assess accurately the improvement.

However, the Royal Air Force has organized field trials on equipments and in particular on an airborne v.h.f. receiver in which the equipment design is such that valves are subjected to very severe vibrations.

This work has been most valuable, because the results obtained in early manufacture permitted the production control levels to be set with reasonable correlation with operational practice. Other than with the earliest deliveries there have been no zero-hour failures. The assessment to date is that in this onerous equipment the trustworthy valves are at least 10 times better than their commercial equivalents.

(10) FUTURE TRENDS AND POSSIBILITIES

(10.1) General

It has been shown that radio valves can be designed and manufactured under mass-production conditions such that they will give a very high degree of reliability in service.

It is important that careful consideration be given to what further action can be taken to achieve even higher reliability.

(10.2) The Valveholder Problem

Basic reliability is now limited by the inevitable mechanical incompatibility between the valve and the present valveholder. The reduction in valve sizes and the introduction of the all-glass base have accentuated the difficulties, and interference tolerances between valve-socket positioning and the base pins in themselves cause potentially higher failure percentages than the targets achieved by the valves alone. The sensible precautions of using pin straighteners and wiring jigs are only palliatives.

The logical approach is to dispense with the valveholder in its present form. Apart from its inherent tendency to cause glass faults owing to mechanical strain, the present designs restrict the exploitation of the valve characteristics because the sockets constitute a source of additional inter-electrode capacitance and also impose serious limitations in maximum voltage and altitude ratings. One other serious failing is that oxidation between the pin and socket surfaces causes considerable variation during life and is particularly objectionable on long-life valves. This last trouble has led the Post Office and designers of telephone repeater equipment to adopt a soldering-in technique using a "spider base" welded to the valve pins. It is understandable that such a method has overcome this particular problem, but its many drawbacks must make it but a temporary measure.

(10.3) Future of the "Wired in" Valve

It is the authors' opinion that the fully "wired in" valve with flexible leads will ultimately come into universal use in industrial and Service equipments, the size of valve being chosen on its dissipation requirement. Thus the subminiature (0·40in diameter) will be employed for low dissipation needs, the miniature (0·75in diameter) and the noval (0·875in diameter) being used where better characteristics and higher dissipations are needed. The objections to their use at present are mainly focused on the problems involved in training personnel in soldering them into circuits, and include also the difficulties of replacement. However, evidence is available, or is being accumulated, to show that it is an out-of-date convention to insist on the valve being a plug-in component. One source of such information can be the Post Office, with its Medresco hearing aid where the lives of soldered-in valves of normal domestic reliability have far exceeded the most optimistic estimates. In addition, an associated company—over a period of 7 years involving some 2 million valves—has experienced no difficulty in training people to perform satisfactory soldering-in operations in this hearing-aid equipment.

There are already available various methods of holding "wired in" valves, but it is the authors' opinion that, whilst these are advances on present techniques, there is scope for more original approaches to these valves, regarding them as new tools and conceiving circuit techniques specifically designed to utilize to the full their additional technical advantages. Such approaches will also give major improvements in the reduction of the weight and bulk of electronic equipment.

It is in this manner that the valve manufacturer will be able to secure maximum valve reliability, and, by close collaboration with him, the progressive circuit designer can achieve the desired goal of complete freedom from valve failures in equipment.

(11) ACKNOWLEDGMENTS

Acknowledgment is made to Standard Telephones and Cables, Ltd., and to the Admiralty for permission to make use of the information contained in the paper.

(12) BIBLIOGRAPHY

(1) INGRAM, S. B.: "Recent Electron Tube Developments," *Electrical Engineering*, 1945, **64**, p. 22.
(2) DREYER, J. F.: "Factors determining Industrial Tube Life," *Electronic Industry*, 1945, **4**, pp. 94 and 148.
(3) CAMPBELL, J. H., and DONSBACH, C. G.: "Analysis of Special Electronic Tube Tests," *Transactions of the American Society of Mechanical Engineers*, 1946, **68**, p. 481.
(4) GORHAM, J. E.: "Electron Tubes in World War II," *Proceedings of the Institute of Radio Engineers*, 1947, **35**, p. 295.
(5) COHEN, V. W., and BLOOM, A.: "Microphonism in a Subminiature Triode," *Communications*, 1947, **27**, p. 18; also *Proceedings of the Institute of Radio Engineers*, 1948, **36**, p. 1039.
(6) ROCKETT, F.: "How's and Why's of Aircraft Tubes," *Aviation*, 1947, p. 85.
(7) ROCKETT, F.: "Improvements in Small Tubes," *Electronics*, 1947, **20**, p. 144.
(8) CHAMAGNE and GUYOT: "Microphony in Radio Valves," *Télévis Franc* (No. 24 Supplement Electronique), 1947, p. 7.
(9) GREEN, N. H.: "Miniature Tubes in War and Peace," *RCA Review*, 1947, **8**, p. 331.
(10) MICHAEL, F. R.: "Tube Failure in ENIAC," *Electronics*, 1947, **20**, p. 116.
(11) ANTON, N., and YOUDIN, M.: "High Altitude Tube," *ibid.*, 1948, **21**, p. 95.
(12) CHERRICK, I. L.: "Rugged Electron Tubes," *ibid.*, 1948, **21**, p. 111.
(13) "Ten Thousand Hour Electron Tubes sought for Industrial Use," *Electrical Manufacturing* (*New York*), 1948, **41**, p. 90.
(14) ROWE, E. G., WYKE, R. E. B., and MACRAE, W.: "Power Diodes: Parts I and II," *Electronic Engineering*, 1948, **20**, pp. 214 and 254.
(15) "Radio Valve Practice," *ibid.*, 1948, **20**, p. 321.
(16) COE, J. C.: "Telemetering Guided Missile Performance," *Proceedings of the Institute of Radio Engineers*, 1948, **36**, p. 1404.
(17) YOUNG, N. H.: "Trends in Military Electronic Equipment Design," *Electrical Communication*, 1948, **25**, p. 334.
(18) HAMMOND, R.: "Vibration Measuring and Recording Devices," *Instrument Practice*, 1948, p. 53.

(19) Willson, A. R.: "Vibration Testing of Airplanes," *Electronics*, 1949, **22**, p. 86.
(20) Jupe, J. H.: "Multi-Purpose British Radio Tube," *ibid.*, pp. 156, 160 and 162.
(21) Knight, C. R., and Haase, A. P.: "A Low Noise (Audio) Input Tube," *Radio and Télévis* (News Radio Electronic Supplement), 1949, **12**, pp. 15, 18 and 31.
(22) Feinstein, L.: "Microphonism Investigation," *Sylvania Technologist*, 1949, **2**, p. 9.
(23) Javitz, A. E.: "Impact of Military Requirements on Product Design," *Electrical Manufacturing* (*New York*), 1949, **44**, p. 142.
(24) Bomford, K. D.: "A Survey of Modern Radio Valves: Part 3—Receiving Valves for use below 30 Mc/s," *Post Office Electrical Engineers' Journal*, 1950, **42**, p. 201.
(25) Alpert, N.: "Microphonics Tester for Vacuum Tubes," *Electronics*, 1950, **23**, p. 78.
(26) McElwee, E. M.: "Statistical Evaluation of Life Expectancy of Vacuum Tubes designed for Long Life Operations," *Sylvania Technologist*, 1950, **3**, p. 16; also *Proceedings of the Institute of Radio Engineers*, 1951, **39**, p. 137.
(27) Devey, G. B.: "Reliability in Electronic Equipment," *Proceedings of the Institute of Radio Engineers*, 1950, **38**, p. 344; also *Electrical Manufacturing* (*New York*), 1950, **46**, pp. 107, 182 and 184.
(28) "Achievement of Reliability in Service Radio Equipment," *Proceedings I.E.E.*, 1950, Part III, **97**, p. 207.
(29) McNally, J. O.: "Life Tests of Oxide-Coated Cathodes—An Engineering Report," *Physical Review*, 1950, **78**, p. 356.
(30) Wenzel, J. A., and Waynick, A. H.: "Microphonism in the Dynamically Operated Planer Triode," *Proceedings of the Institute of Radio Engineers*, 1950, **38**, p. 524.
(31) Vigness, I., and Nowak, R. C.: "Streak Photography," *Journal of Applied Physics*, 1950, **21**, p. 445.
(32) Perkins, T. B.: "Present Possibilities of Improved Tube Design," *Proceedings of Symposium on Improved Quality Electronic Components*, sponsored by the American Institute of Electrical Engineers, the Institute of Radio Engineers, and the Radio Manufacturers' Association, 1950, p. 192.
(33) Moe, R. E.: "Industrial Receiving Tubes," *ibid.*, p. 195.
(34) Moe, R. E.: "An Engineering Progress Report on Reliable Tubes," *Proceedings of the Institute of Radio Engineers*, 1953, p. 43.
(35) Weeks, P. T.: "Miniature and Sub-miniature Tubes. Design and Control Factors affecting Life and Reliability," *Proceedings of Symposium on Improved Quality Electronic Components*, sponsored by the American Institute of Electrical Engineers, the Institute of Radio Engineers, and the Radio Manufacturers' Association, 1950, p. 207 (also *Proceedings of the Institute of Radio Engineers*, 1951, **39**, p. 499).
(36) Dolnick, A. L.: "Premium Performance Sub-miniature Tubes," *ibid.*, 1950, p. 212.
(37) Wyman, J. H.: "Improved Vacuum Tubes for Severe Operating Conditions," *ibid.*, p. 218.
(38) Rich, E. S., and Taylor, N. H.: "Component Failure Analysis in Computors," *ibid.*, p. 222.
(39) Lemmens, H. J., Jansen, M. J., and Loosjes, R.: "A New Thermionic Cathode for Heavy Loads," *Philips Technical Revue*, 1950, **11**, p. 341.
(40) Morse, E. K.: "Long Life Tubes for Industry," *Electronics*, 1950, **23**, p. 68.
(41) Morse, E. K.: "Reliable Electron Tubes for Aircraft Communications," *F.M.-T.V.*, 1950, pp. 20 and 34.
(42) Steinberg, E. B.: "How to extend Electron Tube Life," *Electrical Manufacturing* (*New York*), 1950, **46**, p. 108.
(43) Steinberg, E. B.: "Applying Heater Voltage to Electron Tubes," *ibid.*, 1953, **52**, p. 146.
(44) Knowles, D. D.: "Reliability in Electronics," *Westinghouse Engineering*, 1950, **10**, p. 242.
(45) Gage, G.: "High Reliability in Miniature Tubes," *Electronics*, 1950, **23**, p. 66.
(46) Heacock, D. P., and Wissolik, J. A.: "Low Noise Miniature Pentode for Audio Amplifier Service," *Tele-Tech*, 1951, **10**, pp. 31 and 69.
(47) Lewin, G.: "The Design of Robust Shockproof Electronic Valves," *Le Vide*, 1951, **6**, p. 974.
(48) "How Reliable is a Radio Valve?" *Wireless World*, 1951, **57**, p. 91; *Proceedings I.E.E.*, 1951, Part III, **98**, p. 207.
(49) Acheson, M. A., and McElwee, E. M.: "Concerning the Reliability of Electron Tubes," *Sylvania Technologist*, 1951, **4**, p. 38; also *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 1204.
(50) Fahnestock, J. D.: "Military Requirements for Subminiature Tubes," *Electronics*, 1951, **24**, p. 108.
(51) Bibbero, R. J.: "Electron Tube Ratings at Very High Altitudes," *Tele-Tech*, 1951, **10**, p. 42.
(52) Vedder, E. H.: "Improving Industrial Control Design," *Electronics*, 1951, **24**, p. 104.
(53) Eaglesfield, C. C.: "Life of Valves with Oxide-coated Cathodes," *Electrical Communication*, 1951, **28**, p. 95.
(54) Eaglesfield, C. C.: "Vibration Test for Valves," *Wireless Engineering*, 1953, **30**, p. 57.
(55) Haase, A. P., and Fehr, E. B.: "New Vacuum Tube Materials," *Tele-Tech*, 1951, **10**, p. 30.
(56) Brewer, R.: "Radio Valve Life Testing," *Proceedings I.E.E.*, 1951, **98**, Part III, p. 269; also *Wireless World*, 1952, **58**, p. 239.
(57) Crede, C. C.: "Shock Testing of Airborne Electronic Equipment—Parts I and II," *Tele-Tech*, 1951, **7**, pp. 36 and 66; 1951, **10**, pp. 36 and 68.
(58) Crede, C. C.: "How to evaluate Shock Tests," *Machine Design*, 1951, p. 149.
(59) Hill, D. R.: "New Cathode Design improves Tube Reliability," *Electronics*, 1951, **24**, p. 104.
(60) Steen, J. R.: "The JETEC Approach to the Tube Reliability Problem," *Proceedings of the Institute of Radio Engineers*, 1951, **39**, p. 998.
(61) Steen, J. R.: "Electron Tube Quality Evaluation," *Sylvania Technologist*, 1953, **6**, p. 35.
(62) Acheson, M. A.: "Tube Reliability from the Tube Manufacturers' Viewpoint," Institute of Radio Engineers' Fall Meeting, Toronto, 1951. *Sylvania Technologist*, 1952, **5**, p. 34; also *Proceedings of the Institute of Radio Engineers*, 1952, p. 2.
(63) Cusack, F. H.: "Vacuum Tube Reliability," *Western Union Technical Review*, 1951.
(64) Hunt, G. L.: "A Survey of Quality and Reliability Standards in Electronic Valves for Service Equipment," *Journal of the British Institution of Radio Engineers*, 1951, **11**, p. 519 (Discussions p. 540).
(65) Rowe, E. G.: "The Technique of Trustworthy Valves," *ibid.*, 1951, **11**, p. 525; also *Electrical Communication*, 1951, **28**, p. 257.
(66) Rowe, E. G., and Hunt, G. L.: "Discussions," *Journal of the British Institution of Radio Engineers*, 1951, **11**, p. 540.
(67) Rowe, E. G.: "Trustworthy Valves, Greater Reliability achieved by Improved Mechanical Design," *Wireless World*, 1952, **58**, p. 105.

(68) Rowe, E. G.: "Reliable Valves and the User," *ibid.*, 1952, **58**, p. 377.

(69) Rowe, E. G.: "Technique of Trustworthy Valves" (Condensed Reprint), *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 1166.

(70) Bird, R.: "The Measurement of Microphony in Valves," *Electronic Engineering*, 1951, **23**, p. 429.

(71) Babb, Maynard, A.: "Cold Cathode Tubes and their Application to Airborne Electronic Equipment," *Tele-Tech*, 1952, **11**, p. 40.

(72) Goetz, J. A., and Brooke, A. W.: "Electron Tube Experience in Computing Equipment," *Electrical Engineering* (*New York*), 1952, **71**, p. 154.

(73) Whittier, J. R. E.: "Mechanical Considerations affecting Vacuum Tube Reliability—Parts I and II," *Tele-Tech*, 1952, **11**, pp. 42 and 62, and 1952, **11**, pp. 54, 102, 103 and 106.

(74) Wheeler, W. R.: "Sub-miniature Tubes for Mobile Communications," *Tele-Tech*, 1952, **11**, pp. 40, 108, 110, 112 and 114.

(75) Metson, G. H., Wagener, S., Holmes, H. F., and Child, M. R.: "The Life of Oxide Cathodes in Modern Receiving Valves," *Proceedings I.E.E.*, Paper No. 1137 R, June, 1951 [**99**, Part III, p. 69 (Discussion p. 82)].

(76) Espersen, G. A.: "The 'L' Cathode Structure," *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 284.

(77) Holland, D. O., Levy, J. E., and Davis, H. J.: "Loss of Thermionic Emission on Oxide-Coated Cathodes due to Mechanical Shock," *ibid.*, 1952, **40**, p. 587.

(78) "Valve Factory of the Future," *Scope*, 1952, p. 48.

(79) Koch, D. G.: "Increasing Tube Reliability in Industrial Circuits," *Product Engineering*, 1952, **23**, p. 175.

(80) Welch, P.: "Design of a Low Noise Pentode," *Electronics*, 1952, **25**, pp. 194, 198, 200, 202 and 204.

(81) Hammel, H. A.: "Life Testing Reliable Tubes," *ibid.*, pp. 169 and 176.

(82) Jones, W. R.: "Tube Applications for Increased Reliability," *Proceedings of the Institute of Radio Engineers*, 1952, p. 5.

(83) Brown, A. S.: "Discussion of the Complexity and Unreliability of Military Equipment and the need for Simplification and Increased Life," *ibid.*, p. 11.

(84) McGuigan, W. D.: "Maintenance Minimization in Large Electronic Systems," *ibid.*, p. 23.

(85) Packard, A. C., and Weller, R.: "The Reliability Problems in Guided Missile Development," *ibid.*, p. 34.

(86) Summerlin, W. T.: "Application Engineering for Improved Electronic Reliability in Guided Missiles," *ibid.*, p. 41.

(87) Steinberg, E. B., and Oliver, F. J.: "Sub-miniature Electron Tubes are Reliable," *Electrical Manufacturing* (*New York*), 1952, **50**, p. 84.

(88) "Army Communications," *Wireless World*, 1952, **58**, p. 353.

(89) Knight, C. R., and Harding, K. C.: "General Considerations in Regard to Specifications for Reliable Tubes," *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 1207.

(90) Peddle, E. A. R.: "Electronic Switching," *Wireless World*, 1952, **58**, pp. 421 and 465.

(91) "Production Techniques. Walk in Test Room," *Electronics*, 1952, **25**, p. 248.

(92) Levy, I. E.: "Shielding and Mounting Effect on Tube Bulb Temperature," *Tele-Tech*, 1953, **12**, Section 1, p. 72.

(93) Levi, R.: "New Dispenser Type Thermionic Cathode," *Journal of Applied Physics*, 1953, **24**, p. 233.

(94) Wood, R.: "Reliability of Filamentary Sub-miniature Tubes," *Tele-Tech and Electronic Industries*, 1953, **12**, pp. 78, 112 and 114.

(95) Prager, J. H.: "Performance Evaluation of 'Special Red' Tubes," *Proceedings of the Institute of Radio Engineers*, 1953, p. 8.

(96) Wimpey, E. K.: "Development of Inspection Manual for Reliable Tubes," *ibid.*, p. 22.

(97) Heitner, A. J.: "Adoption of Industry Proposals to Sub-miniature Tube Evaluation," *ibid.*, p. 29.

(98) Bromberg, B. G., and Hill, R. D.: "Reliability of Airborne Electronic Components," *ibid.*, 1953, **41**, p. 513.

(99) Slinkman, R. W.: "Production and Quality Control Aspects of Premium Sub-miniature Tubes," *Sylvania Technologist*, 1953, **6**, p. 43.

(100) Rowe, E. G., and Welch, P.: "The Development of Trustworthy-Valve Techniques," *Electrical Communication*, 1954, **31**, p. 172.

(101) "The Use of Electronic Valves," British Standard Code of Practice, CP 1005, Parts 1 and 2, 1954.

(102) Handley, P. A., and Welch, P.: "Valve Noise produced by Electrode Movement," *Proceedings of the Institute of Radio Engineers*, 1954, **42**, p. 565; *Electrical Communication*, March, 1954, p. 38.

(13) APPENDIX: THEORETICAL CONSIDERATIONS LEADING TO AN EXPONENTIAL DECAY-TYPE FAILURE CURVE

(13.1) Catastrophic Failures

A theoretical treatment of the survival curve of valves is obtainable by the treatment of valve failures in a finite population of valves as a stochastic process. If we denote by $p(r, t)$ the probability that r valves operate at time t, then using only the fundamental addition and multiplication laws of probability, and writing μdt for the probability that a valve fails in an element of time dt, we have the equation

$$p(r, t + dt) = p(r + 1, t)r + 1\mu dt + p(r, t)(1 - r\mu dt) \quad . \quad (1)$$

Eqn. (1) indicates that the probability of r valves operating at time $t + dt$ is the probability that $r + 1$ operated at time t, multiplied by the probability of one failure, plus the probability that r were operating at time t, multiplied by the probability of none failing.

Rearranging eqn. (1) and letting dt tend to zero, we obtain

$$\frac{dp(r)}{dt} = \mu[r + 1p(r + 1) - rp(r)] \quad . \quad . \quad . \quad (2)$$

Solving eqn. (2) by introducing the probability generating function

$$g(t, x) = \sum_{r=0}^{r=n} p(r)x^r$$

multiplying eqn. (2) by x^r, and summing with respect to r we obtain

$$\frac{\partial g}{\partial t} = \mu(1 - x)\frac{\partial g}{\partial x} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

If eqn. (3) is solved by the usual Lagrange subsidiary equations, and the boundary condition that $g(o, x) = x^n$ is applied,

then $g = [1 - \exp(-\mu t)(1 - x)]^n$ if μ is a constant.

The boundary condition states that n valves operate at zero time. Hence the mean number of valves surviving, $\bar{n}$, is

$$\bar{n} = n \exp(-\mu t) \quad . \quad . \quad . \quad . \quad . \quad (4)$$

This shows that if the "death rate," μ, is constant, the expected number of failures is given by eqn. (5):

$$\text{Number failed} = n[1 - \exp(-\mu t)] \quad . \quad . \quad . \quad (5)$$

(13.2) Characteristic Decay

A consideration of the decay of characteristic may be made along similar lines. Consider a characteristic R, and let it take the value r at time t. Then, if for the convenience of the subsequent analysis we imagine increases and decreases in R quantized with probabilities $r\alpha dt$ (the probability of unit increase) and $r\pi dt$ (the probability of unit decrease), and if $q(r, t)$ is the probability of R being equal to r at time t, we may write

$$q(r, t + dt) = q(r - 1, t)\alpha r - 1dt + q(r + 1, t)r + 1\pi dt + q(r, t)(1 - r\alpha dt - r\pi dt) \quad . \quad (6)$$

which, when rearranged, and letting dt tend to zero, becomes

$$\frac{\partial q}{\partial t} = (r - 1)\alpha q(r - 1) - r(\alpha + \pi)q(r) + (r + 1)\pi q(r + 1) \quad . \quad (7)$$

Letting $H(t, x) = \sum_r q(r)x^r$ and multiplying eqn. (7) by x^r, and summing with respect to r, we obtain

$$\frac{dH}{dt} = [\alpha x^2 - (\alpha + \pi)x + \pi]\frac{dH}{dx} \quad . \quad . \quad . \quad (8)$$

Solving eqn. (8) we obtain, if α and π are constants,

$$\text{Mean value of } R \text{ at time } t = R\varepsilon^{(\alpha - \pi)t} \quad . \quad . \quad (9)$$

where R is the value at time $t = 0$.

The results of static life-tests, which are regularly performed on valves on a sample basis, give an indication that this theory is a good first approximation to what actually happens.

The solution of eqn. (8) for H, and the subsequent expansion of H in powers of x will give the function $q(r, t)$—the probability that the parameter will take the value r at time t. This will give, for an equipment designed to close tolerances, the probability of successful operation for a given time, and hence an evaluation of the reliability of the apparatus.

DISCUSSION BEFORE THE RADIO SECTION, 12TH JANUARY, 1955

Dr. J. Thomson: It has been very well demonstrated in the paper that the problem of obtaining valves of improved quality is quite a complex one. Unfortunately each factor in the solution influences every other factor. It is in the complexity of the problem that the chief difficulty lies. If attention is paid only to one factor, a comparatively simple development is possible, and this rather naïve approach to the problem has been made from time to time.

The authors have educed by implication what they consider to be the controlling factors in obtaining reliable valves. My personal view on these is to express the complexity of the problem by stating the conditions. Attention must be paid to the materials employed and to the accurate control of their quality, to the detailed structural engineering of the valve—including, of course, the envelope—to the electrical loading in relation to the electrical potentialities of the structure, to the processing of the structure after assembly, to the ambient conditions in which manufacture takes place, and to the mechanization of the manufacture to the greatest possible extent.

I do not propose to give a supplementary lecture upon these, but it is perhaps worth while to direct attention to the fact that the greatest success will be achieved by the engineer who scores the highest aggregate mark in relation to all of them. But accurate control of quality militates against mechanization. Good structural engineering becomes more and more difficult in assembly as the ambient conditions are improved. Indeed, nothing can be done to improve the quality by attention to one of these factors which will not influence all the others to some degree.

It is actually possible to take the argument one stage further and to add two other factors which are obviously related to it, although at first sight they appear to have little to do with the problem. Mr. Rowe has made the point in introducing the paper. First of all, the greatest improvement in valve reliability cannot be achieved until a production run of the order of one million is undertaken. Secondly, although valves of improved quality must be more expensive in early production than ordinary valves, subsequent large-scale manufacture will give reducing cost with increasing reliability.

The authors have made a very careful study of their parameters over an extended period. The results quoted in the paper are of considerable value to all who are working in this field or using the product of this work. The analysis of the different categories of reliability which they give in their paper is well worth careful study. For despite the intensive and highly successful work described, it is still as true as ever that the common valve-type suitable for long-life applications and simultaneously resistant to shock is not only non-existent but does not appear to be needed.

By this I mean that reliability is not a common factor. A valve can be reliable in this, that or some other way, but there is no method of combining all of them in one.

There is one aspect of this subject of which I should like to make special mention. Owing to the excellent competition between the various home radio- and television-receiver manufacturers, there is a very powerful incentive towards any development which will improve the electrical characteristics of receiving valves. But any programme which is designed to improve the mechanical reliability of a valve must retard the improvements of its electrical qualities. Indeed, it is possible to imagine, in the limit, that a degradation in the electrical properties would be necessary to achieve the required reliability. It follows, therefore, that shallow thinkers will always claim that valves of improved quality have poor characteristics. Their more enlightened brethren will realize that a compromise between electrical and mechanical characteristics is essential.

Dr. G. H. Metson: In Section 6.1.3 the authors take a rather too complacent view of the interface resistance problem. It may or may not be true that American valves are more seriously affected than their British counterparts, but the problem is nevertheless a very real one here. It has been suggested recently that, if interface resistance growth is taken out of British valves, service time may be doubled—a weighty matter if only 10 million valves a year are involved.

In Section 5.2 the authors state that on mechanical shock the anode current of a valve has a general tendency to decrease. With shocks of the order of $20g$ it has been consistently noted at Dollis Hill that the anode current, mutual conductance and total emission normally increase. This phenomenon is always associated with a low level of total emission and is apparently a common feature in indirectly-heated receiving valves. In my opinion the phenomenon is important and should receive serious attention. Have the authors any comments on this?

Air Marshal Sir R. Owen Jones: I am a user of valves, a very considerable user. As Royal Air Force Controller of Engineering and Equipment, one of my great responsibilities is to see that

the Service gets every possible piece of reliable equipment it can. I am very grateful that not only this paper but this large audience shows the interest being taken in improving the reliability of that component of an aircraft which, to be perfectly frank, has in the past been the least reliable.

It is easy to understand how the reliability of valves can affect aircraft and how much more they can affect reliability to-day than in years gone by. You are all perfectly aware that a modern bomber has valves all over it. You can hardly stumble round for valves and the boxes containing them. Moreover, the whole safety and efficiency not only of bombers but of other aircraft depends on valves. Whilst they are unreliable, we just cannot get anywhere effectively.

A deterrent bomber force is an expensive force, but the value of the force is low if it is not reliable. It is particularly low—even lower—if the enemy realizes that it is not reliable. The fighter has to be reliable. It has not the same deterrent effect if the enemy knows that it is not reliable. Again I come back to the point that the valve in the past has been an unreliable component.

Now and in the future we are going to use the valve more and more. Valves play a very important part in navigation to the target. You are all perfectly well aware of this. They play an important part in the release of the bomb and the safe return of the aircraft. Navigation to target and release of bomb are vital in war. Return is extremely important also in peace, lest we lose a portion of our force, through not getting aircraft back from practice and preparations. I have said this before in this hall, and I am not afraid of saying it again. But I am most encouraged by the work done on this paper.

It is extremely valuable, and I do hope it will go on and on until reliability is as high as is humanly possible. So much depends on it from my point of view, and I refer not only to getting the aircraft there and back. When any component is not reliable, the reaction comes round to us in many different ways. If apparatus becomes unserviceable, it has to be put right between flights. The requirement for putting it right is skilled men—what we call advanced tradesmen. There are not many of them. They are not easy to obtain and train up to the required standards, and to have on the spot every time the trouble occurs. The more reliable we can make the equipment, the less we shall have to do in that direction—a most important matter from the point of view of the Service and the country.

I was very intrigued to see from the paper how the tests have been done, but there is one further point I should like to mention. The Services, being cognisant of this reliability question, have been trying to help as best they may. The tests and results shown in the paper are most interesting, but we can add and have been adding something to them. Laboratory tests tell you quite a lot, in fact a very great deal. But we feel actual service tests tell you even more but they are difficult tests to make. It takes a long time to run up a really good total of hours on a component, particularly a valve. It is very difficult indeed under our conditions of working to keep the history of these valves carefully, so that the data obtained shall be reliable and accurate. We have been doing this over the past two or three years to the best of our ability. It requires a lot of men to do it. It requires quite a deal of interference with the work that is going on in the squadrons to keep a history of each little valve and make sure that it is accurate—to remove it when it fails and determine what caused it to fail, and finally to get together the information derived therefrom. We have been doing that to the utmost of our ability, and we shall continue to do it and to collaborate with the Ministry of Supply and the industry to our common end.

I am not going to give details of that. One of my staff is here this evening, and he may be given an opportunity of telling you something about it.

There is one other point. Early on, a Table was put on the screen showing the relative reliabilities of the different components. The second on the list was resistors. I think the figure was 8·8%. Perhaps Mr. Collcutt will confirm that we rate the failure rates of resistors under service conditions as somewhat higher than that in relation to the other components. Perhaps I am going a little far. But I can see from the evidence in the paper that the valve reliability question is well under way, and I do not think it will lose its impetus. Might I suggest that we now direct attention to the other components, starting with resistors?

Mr. R. H. Collcutt: Two types of the reliable valves mentioned in the paper (namely CV131 and CV138) have been in service with the Royal Air Force, and the Scientific Adviser's Department of the Air Ministry has had samples of a thousand of each from early production and another thousand of each later on.

They were fitted in the standard v.h.f. transmitter-receiver in use in the Service, as were also, at the same time, similar samples of normal-production valves as a datum from which to measure any difference in failure rate. The aircraft were not flown particularly to test the valves, but the opportunity of normal work was taken to accumulate figures of their behaviour with age. The accumulation of age, in running hours, was rather slow, and so far it has only been possible to get up to about 500 hours.

At installation of the CV131 reliable valve there were 0·37% failures compared with 11·6% with the normal valve. In use, it was found that there were some good and some bad positions. In the best position with the reliable valve in the first 80 hours of life the failure rate was 0·25 per thousand valves per hour of running which settled down subsequently to a failure rate of 0·1 per thousand valves per hour of running. This compared with 1·1 for the first 40 hours and a subsequent rate of 0·2 for the normal valves. In the worst positions the reliable valve had an initial failure rate in the first 80 hours of 1·7 followed by 1·1, and the normal valve had 10·9 for the first 40 hours followed by 4·6.

The ratios between the good and bad positions were, for the reliable valve, between 6 and 10, and for the normal production valve, between 10 and 20, which compares fairly well with the figure given in Section 8.2.

With the reliable CV138 there were no installation failures, whereas there were 5·5% with the normal-production valve. In service, with the reliable valve the failure rate was 0·2 per thousand valves per hour of running in all positions, whereas with the normal valves there were good and bad positions: the good positions had a steady rate of about 0·8, and the bad positions started off with a rate of 2·6 for the first 40 hours, which dropped to a rate of 0·9. The latter rate then began to increase again at about 0·2 per thousand valves per hour. This was the only B7G valve for which figures are available as evidence in the first 500 hours of an actual increase in failure rate with life. This was not exhibited by the reliable valve.

In assessing the number of failures, particular care was taken to count only those valves whose replacement in service made an unserviceable set serviceable. Nevertheless, when these failed valves were put on a full specification test—both normal and reliable types—it was found that 30% were within the specification limits. It seems that the point has now been reached where, in a modern radar set, the total failures to be expected from valves compares with the total failures to be expected from resistors. Thus, although still further improvements in valve reliability are required, I think the point may now have been reached when, in the Service anyway, one can say that in air-

borne sets valves are no longer going to be the major source of trouble.

Mr. F. M. Walker: The authors describe the work that was carried out in the laboratory on heaters, both helical and coiled. In America, manufacturers appear still to be using the faggot heater, which I believe was at one time used largely by the organization with which the authors are associated. Could they explain why they no longer use this type of heater?

Comments have been made about cathodes and interface growth. It seems to me that there must still be a strong lack of concordance in experience. The authors say they find fairly considerable interface growth with ordinary "O" nickel. As one concerned with valve manufacture I find this distinctly at variance with my own experience, which is that with ordinary "O" nickel interface growth is almost inappreciable under 5 000 hours. It seems to me that there must be variations in valve processing which make this large difference possible.

I endorse the authors' opinion that we require some new material for the insulators inside a valve. I do not think that synthetic mica would remove the fundamental difficulties that we have at the moment, since it would still be somewhat fragile and would still laminate, etc. We should perhaps experience improvements in uniformity and gassiness, but this is not sufficient.

Ceramics have been proposed, but their lack of elastic properties and their somewhat variable dimensional shrinkage during firing make them unattractive.

On stability of valve characteristics, the authors distinguish between long- and short-time stability. I notice that no mention is made of that stabilizing period of 48 hours which is currently in use in America and also in this country. I think the authors were in favour of this method of stabilizing valves some years ago. Could they comment on that point?

No mention is made of rejects, after fatigue tests, due to impaired insulation between electrodes. In my experience, the rejects that we should have expected three years ago after fatigue tests are not the ones we experience to-day. Generally speaking, one finds no disconnections, no internal short-circuits and no glass cracks. Instead, one experiences an occasional insulation or microphonic-noise reject. Perhaps the authors would comment on this.

CV specifications for special-quality valves set a standard of reliability which is probably adequate for most military uses. I think most valve manufacturers would agree that valves can now be made to meet these specifications—not easily, but with increasing efficiency. This may serve to indicate the degree of reliability of special-quality valves at the present time.

Mr. R. Brewer: My experience with special-service CV138 and CV140 valves has given percentage life figures comparable to those given by the authors. In 1 500 special-service CV138's, we had 98·4% life and in 600 CV140's we had 99·4% life. The authors' figure for the latter valve is 100%, but the sample size is not given in the paper. This must have been too small to indicate the true quality of the valve.

I do not think that percentage life is a satisfactory way of indicating the reliability of this class of valve: 99% life sounds so good that we are tempted to overlook the fact that, in the 1 500 special-service CV138's to which I have referred, an equipment using 1 000 such valves might expect to fail once in every 50–100 hours. It is probably better to use the percentage failures or survivors at stated hours rather than percentage life, for the kind of applications which have been considered in the paper. Percentage life is more suitable for trunk communication systems, where long life and valve replacement rates in terms of thousands or tens of thousands of hours determine the economics of the system.

Evidence from the examination of life-test failures points to a confirmation that the survival curve is probably exponential in form for the first few thousand hours of life. At a later time, a much higher failure rate sets in with the commencement of some major characteristic change. The exponential failure curve means that the rate is constant and is not higher in the early part of life, as often stated.

The authors remark that under-loading is a useful procedure for extending valve life. This may be true in some instances, but where ultimate failure is due to the growth of cathode core interface-resistance, reduction in the anode current may worsen rather than improve life.

I agree with the authors on the necessity of having good correlation among valve manufacturers in their testing equipment, and some consideration might be given to the idea of the interchange of valve samples between manufacturers. This has long been the practice in the electric-lamp industry, where it is regarded as an important element in ensuring that lamp makers always speak the same language. As the testing of valves becomes more complex, so does the need for improving the reliability and consistency of the testing equipment.

Mr. M. B. Williams: My remarks relate to the maintenance of trunk communication systems and should therefore be confined to valves in Class 1, i.e. "repeater" valves—rather a nostalgic term nowadays since most of the wideband systems now in service use valves originally designed for radio and television work. It appears from the paper that some useful improvement in service from these valve types could be expected from the use of "special quality" methods of manufacture, pending the arrival of long-life repeater valves with modern performance.

The emphasis in the paper on statistical quality control is welcomed since it assists in the planning of a valve replacement programme, but, to be of full value, it requires long-term consistency in the statistics.

Reference has been made to the valve as contributing largely to failure of equipment. We can consider two kinds of failure, at least in the case of trunk systems. Catastrophic failure of a valve will interrupt the circuit until the unit is replaced. Change in valve characteristics will cause the performance of the unit to go beyond the performance limits but may not at that stage degrade service. Recent analysis on a wideband link showed that complete failure was caused more often by resistors and capacitors than by valves. However, the maintenance effort expended in locating and clearing performance faults due to valve ageing is very large. To reduce maintenance costs, great stability of characteristics over a long life is the desirable aim.

The authors have stated that on logical grounds the valve-holder is unnecessary. From practical experience we know the valve-holder to be a serious cause of service faults, and because it is difficult to control its manufacture it is better eliminated.

At the present time we are having to replace all valve-holders of a certain type on a coaxial-cable route. The original pattern has given good service, but slight modification resulted in one production run having a very indefinite attachment of the valve-holder to its mounting plate. Consequently, whenever a valve is removed, more often than not the valve-holder is pulled through the panel, bringing with it all the associated wires and components.

It is with such experiences in mind that I fully support the authors in recommending wired-in valves.

Mr. R. E. B. Wyke: The authors have shown clearly how reliable some of the new special-quality valves can be when used under proper conditions. As they also have pointed out, it is unfortunate that only too often they are not used properly.

In a rising and rapidly expanding industry such as ours, we have hundreds of very keen young men entering it every year,

full of bright ideas, but, to my mind, often lacking in basic engineering knowledge, and it is from this that the bulk of our troubles come. The educational establishments could help us a great deal by encouraging, possibly during vocational courses, the spending of a little time in the valve manufacturing firms, where students could learn more of correct valve usage. They might even set an odd examination question on that admirable document, the B.S. Code of Practice for valve use, to which—incidentally—a section on special-quality valves will be added.

I was glad to see the authors' reference in Fig. 8 to the resonance of individual sub-components of a valve above 1 kc/s. It is not realized enough by the designers of shock mountings that these resonances can occur. Although the valve designer may shift them, he cannot eliminate them entirely. It is often said that equipment will have no resonances above 1kc/s, but we still get them and find people testing valves for them. I do wish something could be done about this.

With regard to the future, I too agree emphatically with the authors that we shall not get real valve reliability until we get the wired-in valve. I also wonder whether we should not exploit a little more the multi-element valve. It has been shown that the double triode can be extremely reliable. Now that we are particularly concerned with glass failures and bad handling failures, we might use more multi-element valves and so have fewer pips to be knocked off and less scratchings of the bulb. There is an indication that new requirements for commercial and colour television will mean more multi-element valves in use, and this is a point we might follow up.

The authors say that there are two schools of thought on the siting of assembly units for special-quality valves, but they do not say which particular type they like, although I believe Mr. Rowe was quite emphatic about this in a previous paper. These two schools of thought are, first, that the valve assembly should be put right in the middle of the factory, so that, by example, that would make all the other people assemble valves of equal quality, and secondly, that they should be segregated. I prefer the latter course.

As to the introduction of incentive bonus schemes in special-quality valve units, I should like to ask the authors whether that was done with technical blessing or whether it was brought about by other factors.

It is hinted in the paper that the cost-cutting spirit is entering into the reliable-valve business. I do know that there is some truth in this. Makers of reliable valves have been annoyed to find, after they have done a lot of work, that their customers are asking them to come in on a very highly competitive tendering business. I deplore this, and I should be glad if the authors would state their views on the matter.

Mr. H. R. Reid: The paper puts emphasis on valves giving normal lengths of life under moderate conditions of vibration and shock but unfortunately deals with only one group of valves within this class. Centrimetric types such as magnetrons and klystrons, etc., are not included, and because of the special difficulties involved, they no doubt deserve separate treatment on some future occasion. Their reliability in Service equipment, however, is just as important as with the class of valves covered in the paper.

The authors make a strong point of the necessity to produce large numbers of valves. From Figs. 10 and 12 and Table 2 of the paper, the production process appears to be in control only after some 10 000 to 15 000 valves have been produced. From my knowledge of batches of valves produced for the Services and from the Table showing the results of the mechanical tests on 200 valves, I feel sure that controlled conditions can be reached at a much earlier stage of production. Perhaps the authors would comment on this aspect.

In Section 6.1.1, I think something should have been said about the means of controlling the welding process, since satisfactory welding is so obviously a necessity in reliable valves. I believe that in the United States improved welding is obtained not only by electronic timing, but by feeding hydrogen through a tube round the welding head and on to the site of the weld. Can the authors say whether similar techniques have been tried in this country?

It is interesting to note the authors' view that, to ensure valves operating satisfactorily under conditions of vibration at $2g$, tests should be carried out at $12g$.

The "A" and "B" thermal shock tests have always been viewed with some suspicion, although a specification requirement, and it would be interesting to know more of the basis on which the authors claim that their "T" test is more in keeping with the difficulties experienced in the field.

Figs. 13 and 14 demonstrate clearly how a normal valve production can have a significant percentage outside tolerable limits. Regarding the mutual conductance figure on even reliable valves, if the model area is close to the inspection rejection limit, the bulk of the valves could quite quickly fall outside operating tolerances in service. This leads me to a point which cannot be too often repeated—that reliability has to be built into the product and cannot be inspected into it.

I join with Mr. Walker in drawing attention to the omission of any reference to ageing tests. In America, the big producers age valves for some 40–50 hours for stabilizing the valves and rejecting early life failures. I know that the authors have definite feelings on this matter. Would they please comment?

As a member of one of the Service supply departments, I should like to draw attention to their contribution to reliability. It was realized quite early in the reliability programme that an agreed specification between all Service departments and the industry would be long delayed. Contracts were therefore placed with the industry to normal valve specifications but calling for embodiment, into the valves, of the lessons learned by each individual firm from their reliability research and development. Thus the Services have had delivered to them much more robust and improved types over the past two years. These valves have generally been referred to as "improved."

Capt. G. C. F. Whitaker: I should like briefly to state my own experience with regard to the introduction of wired-in valves into Service equipment.

The whole literature of the subject and the activities of such people as the authors all point to the fact that they will lead to a natural increase in reliability which is what, of course, we are aiming at.

In A.S.R.E., while we are obviously actively considering their use, it is probably true to say that at the present time it is in a somewhat tentative way. There seems to be a certain hesitancy to accept the recommendation of the valve manufacturers. In a recent instance, a project arose likely to be used extensively in the Service, with some outside commercial applications. A contractor inquired whether the use of wired-in valves would be acceptable. It seemed to be an admirable moment to introduce these valves extensively in a job which was likely to see both Service and civilian application, and in short to subject the idea to a full-scale practical field test. This proposition was quite acceptable to higher authority.

When, however, it came to the Design Group—a body of gentlemen remarkably like yourselves—it was found that they were very much against doing anything of the sort. The objections were not so much on any philosophical or logical grounds but rather that they feared that considerable trouble might result in practice.

The difficulty, I think, is that people are not yet sure whether

he very great inconvenience of dealing with these valves if, in 'act, they prove to be not as reliable as is hoped, does not swamp he admitted advantages of the elimination of the valve-holder. My own reaction is favourable.

On the Naval Air side, the Ministry of Supply have introduced vired-in valves in a number of equipments. It would be true to ay that Naval maintenance staffs are at present somewhat erturbed at the prospect.

An earlier speaker said that he would, if he had his way, stop ll maintenance staffs from pulling valves out of their sockets. can only say that he cannot have faced the problem of dealing apidly with a piece of equipment which has failed but must be estored quickly.

It is unfortunately true that a high proportion of equipment rouble is due to valve or socket failure and that a quick change f valves, while failing to diagnose the trouble, does frequently lear the fault.

To sum up, there appears, in my experience, to be as yet a ertain lack of unanimity amongst engineers on the overall dvantages of the wired-in technique, and a corresponding esitancy to use it.

Mr. D. C. Birkinshaw: I have formed the impression that the mall valves used in B.B.C. television cameras and associated quipment are giving very unsatisfactory lives. I have not been ble to gather from the paper what sort of average life should e expected from small and miniature valves such as we use, but have always understood that a minimum life of 5 000 hours is o be expected. This in itself does not seem to me to be very ong, but I am finding that, at Lime Grove, for example, the alves in various groups of equipment are giving average lives f only 4 000, 3 000 and even 2 000 hours.

These lives surely must be unreasonably low, and I am anxious o find the cause. Is it that the valves are overloaded by the ircuit designers? Is it that valves are given unrealistic ratings y their manufacturers? As a further possibility have the tandards of valve manufacture been allowed to fall too low?

Whatever may be the cause of this, the fact is that the money pent on valve replacements in the B.B.C. Television Service is, n my view, far too high, and it is money which would be much etter spent on creative programme material.

I should like to ask the authors whether they approve of these alve lives, and, if not, whether they can throw any light on the auses of such low figures.

Lastly, would it be reasonable to ask manufacturers and/or esigners to try to work to a first target in which it would be nusual for a valve to fail at under 10 000 hours?

Mr. P. Lewis: Until we all agree that we can wire-in all valves nd forget about them, perhaps we may put in a plea for a eliable valve-holder. A holder with multiple spring contacts vould give much more reliable service and tend to carry away nore of the heat developed.

Unfortunately the heat in some miniature valves is extremely reat. Is it really necessary to make the bulbs of modern power alves so small that the working temperature given by the makers s over 200°C? I cannot imagine any valve-holder spring emaining resilient when subjected to high temperatures.

Another point in connection with excessive heat must be the ifficulty of maintaining a high vacuum. Also, is there not the ossibility of the electrodes distorting and causing changes f capacitance? This would be very inconvenient in u.h.f. vork since it would upset designs for automatic frequency tabilization.

Glass cracking at the pins is generally blamed on failure to se wiring jigs, but has been known to occur with batches of alves even if jigs were employed.

Is there perhaps some connection between rate of interface growth and atmospheric pollution (notably SO_2) affecting cathode materials before assembly?

It has been suggested that it might be better to run valves at low anode voltages to improve useful life. Our designer has used several miniature types as electrometer valves and has noticed changes of characteristics in practice when used in this application and as d.c. amplifiers, although there can be no question of over-running.

He has also noticed changes of characteristics which appear to indicate that supporting the micas against the glass bulb without having the electrode structure sufficiently rigid within itself can cause shifts in the system when only slight pressure is applied to the glass. The spring of the valve retainer has been known to cause this effect.

There has been quite a lot of trouble with heater-cathode insulation in diodes used in noise-limiter circuits, and especially often in cathode-ray tubes. Industrial television is becoming more and more useful in observing dangerous processes in factories, and reliability of heater-cathode insulation is a point one would wish to see improved.

A most elaborate test set has been devised for checking microphonic and mechanical noise. I have not yet come across the Taft–Pierce fly-weight high-impact shock machine, but it should be simple for valve manufacturers to issue a specification enabling even smaller firms to test valves for noise under identical conditions. A pocket edition of the test set shown in Fig. 14 of Reference 65 is required, made to be slipped over the valve *in situ* and having a pendulum suspended at a certain height, containing a defined mass, rubber-coated to a certain thickness. Deflecting it to a certain angle and letting it fall will give a standard shock to the valve: a great improvement over the old screwdriver-test so rightly deprecated by the authors. Perhaps manufacturers could get together and issue a specification for a small device like that which any firm can make and use without having to invest in a vibration machine and a 100-watt amplifier.

Mr. G. I. Hitchcox: The authors, and the speakers in the discussion, have emphasized that, with special-quality valves, wired-in or baseless connection is both practical and desirable. Here it may be difficult to convince the industrial customer, often still suspicious of electronic equipment and often (e.g. in the very important chemical industry) without the facilities or service staff to carry out a fairly difficult emergency soldering operation.

There is a good case for retaining the octal base for industrial valves, perhaps with a robust metal outer envelope, an important psychological point when selling reliability. The current glass bases at their best have poor connection characteristics, are easily damaged, are not self-locating, provide insufficient mechanical retention, and encourage a construction which makes adequate envelope ventilation difficult. Furthermore, they look fragile and unreliable and thereby prolong distrust.

Mr. F. W. Irons: Although I speak as a valve-holder manufacturer, I strongly support the authors' suggestion for wired-in valves, because this would relieve me of the responsibility of providing valve-holders which both clean the valve pins and make contact with them. Much work is being done on the design and development of special holders for wired-in valves, with particular attention to the dissipation of the heat produced in the valve so that it does not "gas up."

I would mention that there are specifications for valve-holders just as there are specifications for valves. Such specifications are produced jointly by the Services and the Post Office and also by the British Standards Institution. The authors have demonstrated that if we pay a little more for valves we can have better ones. The same applies to valve-holders.

Mr. L. R. Mullin: I should like to speak from the point of view of the equipment designer.

In Class (ii) the authors speak of moderate conditions of vibration and shock to give a normal life for civil aviation. What do they consider a normal life, bearing in mind that aircraft fly for as much as 4 000 hours in one year? We need a very long life, and with equipments of varying utilization in one aeroplane, it is difficult to keep accurate records of how long any one of them has been used. Ultimately, the only way, having reduced initial or early failures, is to produce valves with a very long life in order to get trouble-free operation in a civil airline.

Several people have spoken about what designers do to the valves. I would suggest that it would be much easier for the designer if the individual manufacturer was more honest in the way he rated his valves. The 6AQ5 is rated at 12 watts. Valve manufacturers specify a maximum of 230–250°C bulb temperature. They do not tell you that if it is run at 12 watts at normal room temperature it will rise to 250° C.

It is generally assumed that valves in the preferred list are suitable for running in an aircraft, and the heaters can be run in series at 19 or 28 volts, as the case may be. One very important point is often overlooked. The thermal capacities of the heaters must be very much alike if they are to be run in series. CV140 is an example. If it is used in series with other valves and it happens to be a single chain, the heater will almost certainly blow after a very few switching cycles.

Another effect which is not quite so obvious is the effect of a small over-voltage surge on high-mutual-conductance valves and subminiature diodes. The surge frequently causes active material from the cathode to be deposited on the grid, with consequent grid emission troubles. In the subminiature diode a low back-resistance results.

We all hope the improved valves will give longer service, and as designers we shall endeavour to use them in the right manner. But I would suggest that, for civil aviation, the wrong types are being standardized, and the few that are correct are given the wrong labels. In civil aviation, equipment can be sold only if it uses the so-called ARINC range of valves; this applies not only in the United States but also in Europe and Australia. We are having to import valves at the moment in order to export equipments, and it is hard enough to sell equipments without having to explain that a QD77, CV4007, or any other of the many equivalent British types, is really a 5726. So for civil aviation, will valve makers making valves to the same specification please give them the same internationally known label?

Mr. R. E. Clark (*communicated*): About a year ago I was privileged to discuss the whole problem of reliability of electronic equipment in the United States with many of the leading people in the Services and in industry, including ARINC. I feel sure we are not behind the Americans in valve development and reliability: but I rather think we are behind in production of reliable resistors. This view is based partly on a very thorough examination of some United States wire-wound resistors, first by X-ray and then by dissection, followed by a most careful analysis of all the materials involved. This work was undertaken in the Admiralty Materials Laboratory.

On the subject of co-operation, I was most impressed by the enthusiasm of the Advisory Group on Reliability of Electronic Equipment (with the happy short title of AGREE) sponsored at the highest level by the United States Department of Defence. Parallel with this Service Group, the Radio, Electronic and Television Manufacturers Association has formed its own enthusiastic committee on reliability in Service equipment. These two bodies, co-ordinating inside the Services and industry respectively, collaborate closely in all problems involved in providing the user with reliable equipment.

In this country the machinery exists, and on the whole works well, for co-operation between the Services and industry through C.V.D. in the case of valves, and through R.C.R.D.C. and R.C.S.C. in the case of other components; but it is my experience that no such machinery exists inside or outside the Services for co-operation towards producing a common philosophy for design and maintenance of complete equipment. It is my conviction that this is needed and can bring about a great simplification of the whole problem. It appears to me to be a necessary step which must be taken if full benefit is to be obtained from the work of the authors and others on the components of an equipment.

Mr. C. E. Clinch (*communicated*): In the field it is necessary to define the end of a valve's life as the time when the valve will no longer satisfactorily operate the panel into which it has been connected. This definition is not ideal from some aspects of investigations into longer-life valves, but since the field staff's prime function is to keep the service going, it is not possible to keep a valve in service just because its performance is satisfactory on a valve tester if changing the valve enables service to be restored.

It has been stated that 30% of the valves rejected in a field trial were satisfactory on re-test although rejected by the normal maintenance staff. A similar result was obtained by the Post Office in a field trial about 7 years ago, but special analysis of these failures gave the following information:

(*a*) Many valves classified as satisfactory on re-test were rejected by the field staff for noise or microphony, an aspect not checked on the re-test.

(*b*) Some valves were correct at specification voltages but found to be outside limits at the lower voltages actually used in the equipment.

(*c*) Some faults reappeared only after long warming-up periods, and these were not found on the first re-test.

(*d*) Some faults were never proved but may have been cleared by the action of cooling down the valve when it was removed from service.

It is not right to blame the field staff for being too ready to change valves unless all these aspects have been checked.

Also, it has been stated that CV specifications form the manufacturer's contract for the supply of the valves to an agreed standard, and they form the limits for the design and maintenance engineer. It is, of course, fundamentally wrong that the design and maintenance engineer should work to the same limits as the manufacturer, although this has been done. If, for maintenance, use is made of the manufacturer's specification, some valves will be outside limits in a few hours.

It will be interesting to see actual survival curves for valves of improved quality. The plots of existing-type valves do not quite conform to the exponential decay curves given by the authors and in earlier papers,* especially over the first few hundreds of hours. Unfortunately this is the period during which most manufacturers test their valves. It is hoped that the agreement will be sufficiently close with improved-quality valves to make it possible for the formula derived by Campbell† for electric-lamp replacements to be used as a guide to the economics of valve replacements.

Much work has been done on the small receiver-type valves, and there has been a suggestion that resistors and capacitors should be the next to be investigated to achieve greater reliability, but the work on valves is not complete; much work is still required on the more expensive valves for u.h.f. operation including the travelling-wave amplifier.

* E.g.: Lewis, N. W.: "Notes on Exponential Distribution in Statistics," *Post Office Electrical Engineers' Journal*, **41**, Part 1.
† Campbell: *Journal of the Royal Statistical Society*, 1941, 7, No. 2.

THE AUTHORS' REPLY TO THE ABOVE DISCUSSION

Messrs. E. G. Rowe, P. Welch and **W. W. Wright** (*in reply*): Our particular thanks are due to Air Marshal Jones for his kind comments and to his organization for providing so much valuable field correlation information. The figures quoted by Mr. Collcutt demonstrate the big advances that have been achieved. It is our hope that adequate action will be taken now with other components to ensure that maximum *equipment* reliability is secured.

Mr. Walker and Mr. Reid both ask about our views on 48-hour stabilizing periods. We consider that stabilization is better achieved by more careful attention to manufacturing processes, controlled by emission checking before and after a [illegible]-hour life run.

We agree with Mr. Brewer's remarks regarding better methods of expressing life figures, but those quoted were in definite relationship to published figures in the United States. We would also welcome closer co-operation on standardization of test equipment.

Many speakers supported the wiring-in of valves, pointing out the shortcomings of present valveholders. At the same time they have stressed the nervousness of the user to be committed to such techniques because of possible maintenance troubles. It is a case of a dog having a bad name, but surely there has been ample evidence from the user in these discussions to dispel such fears. Future effort on improved reliability depends on the acceptance on wiring-in as a policy. Notice should be taken of comments from Messrs. Reid and Clinch on special types—such as centrimetric and small transmitting types. The reliability of equipment depends on the individual reliability of all components, and much work should be done on these types which can provide many difficult problems.

Mr. Wyke's point regarding customer pressure for reduced prices for reliable valves is well made. It is vital that the customer should pay attention to the risks at stake, particularly in Service equipment and passenger transport, where safety and lives are of paramount importance. Further improvements in reliability will be increasingly expensive, and in many cases they will be achieved by refinements of design and improvements introduced on the production line; and we admire the American approach whereby the prices paid permit the manufacturer steadily to improve his product beyond the target set by existing specifications.

With reference to Mr. Reid's comment on our emphasis on quantity manufacture, reliability stems from a satisfactory valve design made under satisfactory manufacturing conditions. Even small quantities of a satisfactory design can have good reliability, but it is all a matter of the standard of reliability demanded. His comment on welding is also very important, and there is extensive work proceeding to get this process under highly controlled conditions. The evolution of the "T" test for glass testing has been from large-scale analysis of field returns whose failure pattern could not be simulated by "A" or "B" testing.

Mr. Mullins asks about life expectancy. The Service demand was centred on a 1 000-hour period, but it is fortunate that all the evidence on the work done to date shows that much longer lives will be normal. The rating of valves under widely varying ambient conditions is being given more attention, and it is hoped to publish more information on this in the near future.

Our answer to Mr. Birkinshaw's complaint about small valves giving short lives in his equipment is simply that there is ample evidence that the same valves are satisfactory in a multitude of other equipments. By implication this leads us to the importance of close liaison between circuit designers and valve manufacturers' application departments. We have proved repeatedly that we have been able to achieve the results desired without subjecting valves to overload, and it is surprising to us that with the reputation of the reliability of their equipment at stake, designers do not make more use of the services available.

The life target proposed is not an unreasonable one and is certainly not beyond the capabilities of mass-produced valves.

Mr. Clark's communicated comments show that moves are in hand in the United States to achieve close co-operation between all sections concerned in electronic equipment. It is vital that similar action should take place in this country. One method of prompting this might be for The Institution to sponsor a convention to embrace all aspects of equipment reliability.

621.317.361.018.422: 621.391: 621.38

Paper No. 1836 M
May 1955

MEAN FREQUENCY DETERMINATION OF NARROW-BAND NOISE SPECTRA

By W. W. H. CLARKE, Ph.D., B.Sc., Associate Member, and R. F. NIKKEL, M.A.Sc., B.A.Sc.

(*The paper was first received* 12*th July*, 1954, *and in revised form* 24*th January*, 1955.)

SUMMARY

There are an increasing number of problems involving the detection or measurement of signals which properly comprise noise, but possess certain features distinguishing them physically from homogeneous random noise. The paper deals with the measurement of mean frequency for such a signal whose energy/frequency distribution approximates to a symmetrical Gaussian distribution. The generation of synthetic signal sources for controlled experiments and the design of optimum measuring equipment are described and illustrated by experimental results using a magnetic-amplifier servo loop for automatic measurement. The effect of the signal deviating from its assumed characteristics is also discussed, and it is shown that the measuring equipment described can tolerate a certain amount of signal asymmetry without undue loss of accuracy.

LIST OF SYMBOLS

W_s = Energy in signal.
$dW_s(f)$ = Signal energy (a function of frequency) in bandwidth df.
W_n = Energy in noise.
$dW_n(f)$ = Noise energy (a function of frequency) in bandwidth df.
f = Frequency.
f_0 = Mean frequency of spectrum.
K = Constant of proportionality in signal-energy equation.
$a = b/f_0$ = Gaussian exponential coefficient.
b = Constant.
B = Constant of proportionality in noise equation, or a function of frequency.
f_m = Frequency of symmetry of a pair of sampling filters.
δ = Frequency error $(f_m - f_0)$.
$\sqrt{\overline{V^2}}$ = R.M.S. deviation of the rectified output of one filter from the mean rectified potential.
V' = Mean rectified potential.
CR = Time-constant of rectifier.
f_h = Width of filter pass band.
W_1 = Energy output from filter 1.
W_2 = Energy output from filter 2.
V_1 = Rectified voltage output from filter 1.
V_2 = Rectified voltage output from filter 2.
f_N = Bandwidth of noise.
f_ϕ = Controlled oscillator frequency.
c = Constant in noise distribution equation, $B = c/f$.

(1) INTRODUCTION

Most problems involving a.c. signals are clear cut as regards the nature of the signals and the unwanted background which is called "noise." The signals generally take the form of one or more sine waves, or at least can approximate closely enough to such waves for their band spread to be ignored both in interpretation and in accuracy. A different class of problem occurs when it is desired to make use of a signal which is composed neither of sine waves nor of random (white) noise, but has some of the characteristics of each.[1] Such problems arise when it is desired to measure small increases in noise level, in particular spatial directions or over certain regions of the spectrum,[2,3] or where it is required to extract particular components from noise spectra.[4]

The paper is concerned with the related problem of measuring the mean frequency of an energy/frequency distribution which is too wide to be investigated by such techniques as axis-crossing counters or synchronous motors, and which may be accompanied by a substantial level of wide-band noise. Theoretical and experimental work has been done, particularly on the measurement of a Gaussian distribution of energy accompanied by level white noise, and the technique described is believed to represent an optimum for the type of measurement required, even when the spectrum deviates from a Gaussian distribution.

The paper deals with the engineering approach to providing a system which can continuously monitor the mean frequency of a spectrum whose width is of the order of one-fifth of the mean frequency, in the presence of noise. It bears little relationship to the possible laboratory problem of investigating a given distribution for which there is a multiplicity of standard equipment.

(2) THE PROBLEM

Experimental and theoretical work relates particularly to an energy/frequency distribution defined by a symmetrical Gaussian equation as follows:

$$dW_s(f) = K\exp\left[-(b/f_0)^2(f-f_0)^2\right]df \quad . \quad . \quad . \quad (1)$$

where $b = 10$; giving a width between $(2\cdot71828\ldots)^{-1}$ levels of $0\cdot2f_0$.

This is assumed to be accompanied by noise as follows:

$$dW_n(f) = Bdf \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

The case where B is constant has been principally considered, and is the only case investigated experimentally. The case $B \propto 1/f$ is considered theoretically.

Actual values for f_0, the mean frequency, range from 1 to 7 kc/s and the noise lies between 100 c/s and 8 kc/s, eqn. (2) being invalid outside this range. This large range of 7 : 1 for the Gaussian spectrum is the significant factor in determining the parameters of the measuring apparatus. The experimental work described employed synthetic sources for these signals, and they are discussed as being complementary to the special equipment developed.

(3) THEORETICAL APPROACH

It is convenient to appreciate the problem first from the assumption that the energy in any element of bandwidth may be selected and used as a contribution to a direct potential, and that the energy in a chosen bandwidth may be compared in this way with that in another. The assumption of symmetry leads naturally to the idea of comparing signal level at equal deviations on either side of a chosen "trial value" for the mean frequency. This is inevitably the basis of any measuring technique suitable for the problem, and Fig. 1 shows the extent of the available error signal for a trial value f_m which deviates from the mid-frequency f_0 by an amount δ. The energy/frequency distribution

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Dr. Clarke and Mr. Nikkel are at the Defence Research Telecommunications Establishment, Ottawa, Canada.

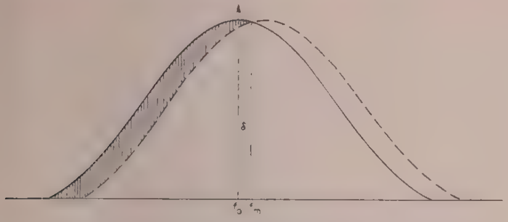


Fig. 1.—Available error signal related to dividing frequency between filters f_m.

Available error signal.

is centred on f_0, and its image, folded about f_m, is shown dotted. The shaded area then indicates the differences between energy contributions from frequencies equally above and below f_m. If these energy differences may be integrated for a chosen interval on either side of f_m, a maximum contribution, $f_m \pm 1/a\sqrt{2}$, is obtained where $a\,(=b/f_0)$ is the exponential coefficient in the spectrum eqn. (1). The total available error signal varies with δ in the manner shown in Fig. 2.

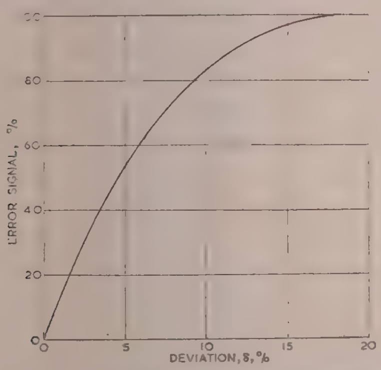


Fig. 2.—Variation of error signal with frequency deviation.

It would seem that the sampling points should have a separation of $(\sqrt{2}/a)$, but of course a varies inversely with f_0; and f_0, in the particular application to which the practical work herein relates, varies over a range of 7 : 1. This wide range associated with a proportionate spectrum width is doubtless a common feature in most similar problems. The circuit complications which would allow sampling points whose separation varied with f_m are considerable, and there are fundamental disadvantages in the narrow-band sampling system. In a narrow-band sampling system the pull-in range of f_m towards f_0 would be proportional to f_0, and at small values of f_0 would represent a very small part of the range of excursion of f_0. The amount of signal used would be a very small part of the available energy, and further amplification would be required. The restricted bandwidth of samples would make small random-error signal variations more likely unless a long time-constant were employed.

The maximum error signal, requiring the shortest rectification time-constant, may be obtained by the use of filters as wide as possible, one with a pass band extending upwards from f_m and the other downwards. This appears to be the ideal arrangement in the absence of unwanted noise, particularly as it may be shown to have advantages when the spectrum deviates from symmetry.

Unfortunately noise is always a feature of practical equipment designed with the minimum necessary power, and there is therefore a restriction on the optimum extent of filters on either side of f_m. The factors influencing the filter bandwidth are the required pull-in range and loss of sensitivity through the presence of excessive noise. Fortunately there is an effective reduction of noise bandwidth when the energy in the two sampling filters is rectified into a long time-constant. If $\sqrt{\overline{V^2}}$ is the r.m.s. deviation of the rectified output from one filter $-V'$ is the mean rectified potential, and $\sqrt{\overline{V^2}}/V'$ will be of the order of $1/CRf$, where CR is the time-constant of the detector and f is the bandwidth of the filter. The r.m.s. error voltage due to noise is $\sqrt{2\overline{V^2}}$. Hence a large noise tolerance should be possible with a suitable choice of time-constant. However, the presence of a large amount of noise in a system employing automatic gain control is inevitably accompanied by a small amount of signal, and hence the true error signal $(V_1 - V_2)$ will be correspondingly small, giving reduced sensitivity. In the practical example, which is the basis of the paper, the energy/frequency spectrum is subject to a good automatic gain control. Hence the energy of the total spectrum, whether at $f_0 = 1$ or 7kc/s, will be the same, and therefore the signal entering filters with bandwidths wide enough to accept most of the energy of the $f_0 = 7$kc/s spectrum will contain as much or more energy at lower values of f_0, and the filters will not receive more noise. It will be shown that the wide range of mean frequencies introduces a limitation on the bandwidth of filters which may be employed. Apart from reasons associated with particular systems, limitation of filter bandwidth is necessary from noise considerations alone. This may be seen from the following, which assumes that the noise occupies a sufficient bandwidth to give a level contribution across the full range of each filter. For the wider filters this implies a restricted range of spectrum mean frequencies.

The noise energy contributes to each filter an energy $W_n(f_h/f_N)$.

The signal contributes to filter 1 an energy W_1, and to filter 2 an energy W_2, where

$$W_1 = \int_{f_m - f_h}^{f_m} K \exp\left[-a^2(f_0 - f)^2\right] df \quad . \quad . \quad . \quad (3)$$

$$W_2 = \int_{f_m}^{f_m + f_h} K \exp\left[-a^2(f_0 - f)^2\right] df \quad . \quad . \quad . \quad (4)$$

$$W_s = \int_{-\infty}^{\infty} K \exp\left[-a^2(f_0 - f)^2\right] df$$

f_h = Bandwidth of each filter

$a = 10/f_0$

Let the r.m.s. voltage from filter 1 be V_1, and the r.m.s. voltage from filter 2 be V_2; then

$$V_1 \propto \sqrt{\left\{W_n\left(\frac{f_h}{f_N}\right) + \int_{f_m - f_h}^{f_m} K \exp\left[-a^2(f_0 - f)^2\right] df\right\}} \quad . \quad (5)$$

$$V_2 \propto \sqrt{\left\{W_n\left(\frac{f_h}{f_N}\right) + \int_{f_m}^{f_m + f_h} K \exp\left[-a^2(f_0 - f)^2\right] df\right\}} \quad . \quad (6)$$

These equations may be used to calculate the manner of variation of V_1 and V_2 and their difference with f_h, for chosen values of W_n, W_s, f_N, f_m and f_0. It is justifiable to assume for any relatively narrow bandwidth that the rectified signal voltage is proportional to the square root of the energy.

The curves in Fig. 3 show the variation of $(V_1 - V_2)$ with f_h for various values of W_n. In Fig. 3A the error signal is normalized to make $W_s + W_n = 1$, which corresponds to a good

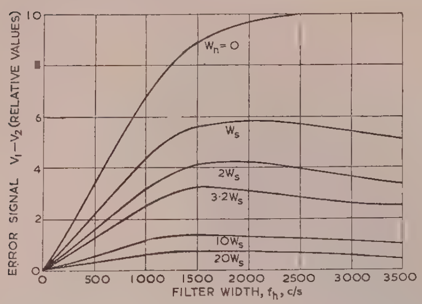


Fig. 3A.—Error signal with various noise levels, normalized to make $W_n + W_s = 1$.

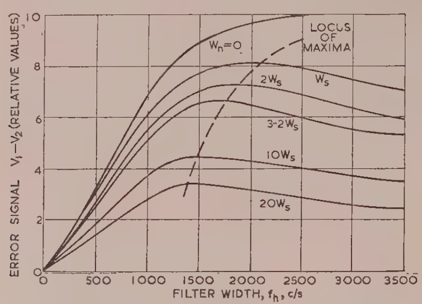


Fig. 3B.—Error signal with various noise levels and constant spectrum energy.

a.g.c. characteristic. In Fig. 3B the signal energy W_s is assumed to be constant, and the locus of maximum error signal is indicated. For the case chosen, $f_N = 9$ kc/s, $f_0 = 7$ kc/s and $f_m = 1{\cdot}01 f_0$, and it may be seen that, for equipment to work in the presence of considerable noise levels, there is no reason to make filters with bandwidth wider than some 1 500 c/s.

(4) EFFECTIVE FREQUENCY SHIFT OF FILTERS AND FILTER RESPONSES

It has been assumed that the filters postulated can be adjusted in frequency without change of shape and that they each pass equal bandwidths in adjacent bands and reject all other frequencies. By mixing the calibrated sine-wave oscillator signal with the spectrum to be measured, it is possible to produce signals at addition and subtraction frequencies which are reproductions of the original spectrum in a higher frequency band. Variation of the calibrated oscillator frequency causes the spectrum to move in the higher band, and if fixed filters are located there, the effect is substantially the same as adjusting the filters in the original band. It is desirable to use a frequency-addition process, so that the intermodulation signals from the different frequency components in the whole spectrum including noise do not fall within the frequency range of the filters. At the same time it is difficult to make filters which have adequate discrimination against the major components of the input to the mixer, which are the controlled oscillator signal and the Gaussian spectrum. Therefore, if these signals are minimized by the use of a balanced mixer, practical filter responses can provide the rest of the required rejection.[5, 6] Fig. 4A indicates the distribution in frequency of the spectra, the controlled oscillator and the filters which work in accordance with the block schematic (Fig. 4B). A magnetic-amplifier servo loop is shown using the error signal ($V_1 - V_2$) to correct the setting of the calibrated controlled oscillator. In the balanced condition, $f_0 + f_\phi = 20{\cdot}5$ kc/s, where f_ϕ is the controlled oscillator frequency.

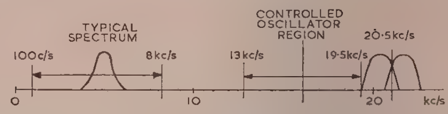


Fig. 4A.—Diagram of measuring process.

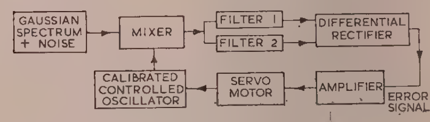


Fig. 4B.—Schematic of measuring circuit.

It is necessary that the pass band should extend by at least the bandwidth of one filter below the mean frequency of the lowest frequency spectrum in order that equal noise levels be presented to both filters, otherwise errors occur at the lower frequencies.

A number of matched pairs of filters have been designed for various arrangements of this kind, both for addition and subtraction systems, and for different bandwidths. Fig. 5 shows

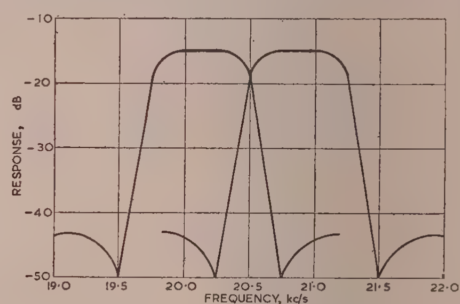


Fig. 5.—Overall response of filter.

the responses of a pair of filters which were designed to work in the addition system illustrated in Fig. 4. It has been found that satisfactory filters can be made using one symmetrical mid-shunt m-derived band-pass section. The most satisfactory balanced mixer that has been tried appears to be the ring modulator.

The practical filter responses are not sharp-sided but have finite slopes, which give some response outside the chosen bandwidth and present a problem regarding the response level chosen for the "crossover frequency."

Particular attention must be paid to ensure that when one filter characteristic is inverted about the crossover frequency the inverted characteristic coincides exactly with the other filter characteristic. The skirts of the characteristics accept noise over the full signal bandwidth, but if 20 dB or more of rejection is available the power contributions of noise at these rejected frequencies will be sufficiently low not to introduce appreciable errors. The symmetry requirement applies to all responses less than 20 dB below the band-pass level.

The slopes of the filter characteristics are most important, because they can contribute to asymmetry with corresponding measurement errors which vary with noise level, and because the sensitivity depends on the sharpness of the characteristics. The slopes on the adjacent sides have practically no effect on the wider spectra, but are the only appreciable effect on the narrower ones. It is possible to compute the effect on the error signals of given slopes of the filter characteristics, but the proper approach becomes one of measurement when the elementary conclusions based on perfect filters are extended by simple

approximate deductions to show that reasonable slopes of the filter responses do not cause great deterioration of the expected performance. For filter responses crossing at 6dB below their band-pass level and falling within a frequency of 100c/s to 6dB below their band-pass level, there is negligible effect on the error signal for $f_0 = 5$kc/s with 1% deviation from f_m; and for $f_0 = 1$kc/s the error signal is more than half that for $f_0 = 5$kc/s.

(5) SYNTHESIS OF SPECTRUM SOURCES

Synthetic Gaussian spectra of fixed frequency have been obtained employing a white noise source passed through two π-sections of a constant-k filter which were adjusted until the response was substantially similar in shape to the required Gaussian response, by altering the Q-factor of the coils by means of series resistance.

A more flexible variable Gaussian signal was also used in carrying out tests on matched pairs of filters. A constant-k filter was used on the frequency 4kc/s, and its output was recorded on a 2-channel tape recorder. At the same time the second channel was used to receive a fixed frequency of 4kc/s. The tape-drive mechanism of the recorder was modified by the installation of a variable-speed d.c. motor, so that any desired spectrum with f_0 in the frequency range 1–6kc/s could be obtained by adjustment of the motor speed, while a monitor of the nominal frequency was available from the second channel. The tape recorder is ideal for the practical case where spectrum width is proportional to the mean frequency. The arrangement is illustrated in Fig. 6, which also indicates a provision for adding a variable quantity of random noise. The energy in the random noise is restricted to that within the total spectrum pass band by the use of an a.f. filter.

The variable synthetic Gaussian source with the inclusion of noise has been used in experiments to be described, for the investigation of the properties of the servo system. The signal/noise ratio of the synthetic signal has already been defined by implication in Section 3 as the ratio of the energy in the Gaussian spectrum to the noise energy (without the Gaussian spectrum) in the total bandwidth. To establish the signal/noise ratio, the Gaussian signal level may be measured at the filter output (Fig. 6) in the absence of noise, using a vacuum-tube volt-

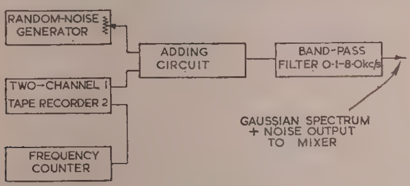


Fig. 6.—Schematic of arrangement for providing synthetic signal with noise.

meter giving an r.m.s. reading. It is then possible to calculate the reading which would be given by the required noise level in the absence of the Gaussian signal and to adjust the noise level to give that reading. For instance, to obtain a unit signal/noise ratio, the noise level is adjusted to give the same voltmeter reading as was obtained with the Gaussian spectrum alone. It is more difficult to obtain an r.m.s. reading for the wide-band noise than for the narrower-band filter outputs, but the very wide range of noise levels makes the possible 0·3dB error in noise-level measurement insignificant.

(6) EXPERIMENTS WITH VARIOUS ARRANGEMENTS

The most important properties of the frequency-monitoring system are the accuracy with which the mean frequency f_0 may be determined and the maximum deviation of f_m from f_0 at which correction is obtained. The former property may be defined as sensitivity and the latter as pull-in range, and both may be measured as a frequency deviation of f_m from f_0 with reference to Fig. 1. According to Fig. 3 it is to be expected that for a given error $(f_0 - f_m)$ the error signal varies with filter bandwidth and noise level. Thus the sensitivity should also vary with these parameters. It may also be deduced that the pull-in range will decrease as the noise level increases. Experiments with various arrangements have been performed in order to measure the relative sensitivity and the pull-in range as functions of noise level. The relative sensitivity is defined as the least deviation of f_m from the value giving zero error signal which will cause the controlled oscillator to start changing its setting. For accurate measurement of sensitivity it was found desirable to use a calibrated variable oscillator in place of the controlled oscillator, but to let the control operate through the servo loop. The deviation of the calibrated oscillator from the null-error-signal position required to make the controlled oscillator adjust continuously, is a measure of sensitivity. The minimum difference between frequencies causing continuous adjustments in reverse directions is twice the sensitivity.

By means of a Gaussian spectrum having a mean frequency between 1 and 6kc/s in a subtraction system, it was found possible to use an arrangement as indicated in Fig. 7. The

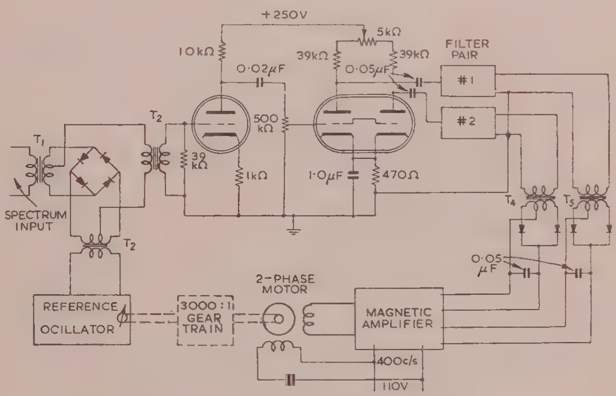


Fig. 7.—Frequency-measuring circuit.

reference oscillator was controlled by the magnetic-amplifier servo loop over a range of frequencies from 8·5 to 13·5kc/s. It was mixed with the Gaussian spectrum to produce a difference frequency which was amplified and passed through a pair of band-pass filters centred on 7·5kc/s and each having 350c/s bandwidth. The two filter outputs were rectified and supplied differentially to two identical control windings of a push-pull magnetic amplifier. If the mean difference frequency was either greater or less than 7·5kc/s one filter passed more energy than the other, with the result that the magnetic amplifier provided an output. This output adjusted the frequency of the controlled oscillator through the control phase of a 2-phase servo motor.

The pull-in range is plotted against signal/noise ratio in Fig. 8 for the different mean-frequency spectra. It may be seen that a unit signal/noise ratio has only a slight effect on the operation and that a signal/noise ratio of -10dB can be tolerated with a reduced pull-in range. The sensitivity was found to be some 20c/s with $f_0 = 2\cdot5$kc/s, and on replacing the Gaussian by a monochromatic signal at 2·5kc/s the sensitivity became 5c/s. The latter figure is a measure of the effect of the slopes of the filter responses near the frequency of symmetry.

This arrangement was not investigated further because the

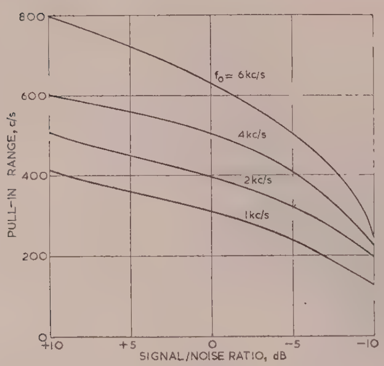


Fig. 8.—Pull-in range using magnetic-amplifier differential.

low impedance of the magnetic-amplifier input windings made it impossible to employ a long time-constant, and the provision of a high-impedance detector followed by an isolating differential amplifier was considered desirable.

A "long-tailed pair" was employed to pass a current to the magnetic amplifier proportional to the difference between the rectified filter outputs. The circuit given in Fig. 9 was used for

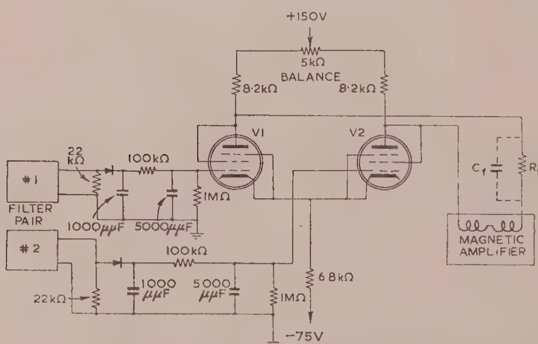


Fig. 9.—"Long-tailed pair" differential circuit.
V_1, V_2: Valve type 6AK5 or equivalent.
R_f: Forcing resistor.

some tests of the arrangement. Since the grid input impedances of the long-tailed pair were high, any suitable time-constants for the rectified filter outputs could be employed. The difference between these signals was found by the long-tailed pair and not by the magnetic amplifier. Thus the sensitivity of the system was improved greatly, and it permitted the use of a forcing resistor in the magnetic amplifier to adjust the transient response. These features were all found to be advantageous. Further advantages were discovered in the facility for adjusting the balance control to offset minor differences in the filter responses, and in the elimination of an output transformer from the circuit.

More accurate measurements could be made using the fixed mean frequency sources, and both pull-in range and sensitivity were measured at 1, 2·5 and 4kc/s.

Comparative measurements were made with a frequency-addition method of measurement involving a pair of filters with responses on each side of 20·5kc/s and with bandwidths of 750c/s as indicated in Figs. 4 and 5. This latter arrangement is suitable for mean spectrum frequencies between 1 and 7kc/s. Figs. 10 and 11 illustrate for the long-tailed-pair circuit the behaviour of pull-in range and sensitivity, respectively. If the curves for the subtraction system are compared with the performance of the magnetic-amplifier differential (Fig. 7), it may

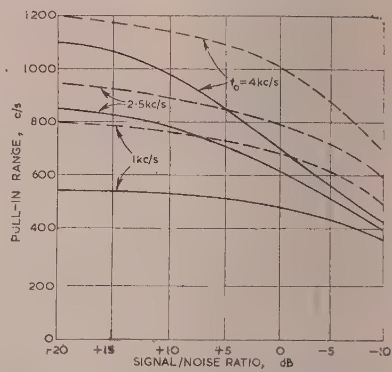


Fig. 10.—Pull-in range using long-tailed pair.
——— Frequency subtraction.
– – – – Frequency addition.

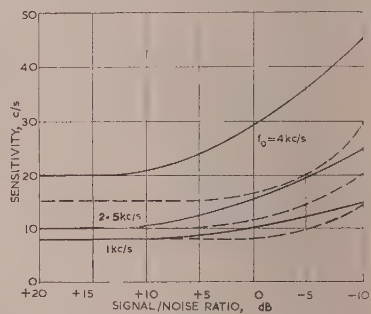


Fig. 11.—Sensitivity using long-tailed pair.
——— Frequency subtraction.
– – – – Frequency addition.

be seen that the long-tailed pair improves both properties of the measuring circuit. With the frequency-addition system, there is a distinct improvement in pull-in range, but it is not so large as might have been expected, bearing in mind the substantially greater bandwidth used in the addition filters. The feature of operation in the presence of large quantities of noise is not appreciably different from the case of subtraction filters using the same error-signal circuits.

A critical feature of the equipment is the time-constant of the detection circuits in Fig. 9. This determines the final waveform of the error signal as applied to the control windings of the magnetic amplifier. If the time-constant is long, the result is an approach to peak rectification and the disappearance of all higher-frequency transients. Too short a time-constant makes inefficient use of the available error signal, which occurs as a series of transient voltages not sufficiently enduring to build up useful currents in the magnetic-amplifier control windings. Direct input to the magnetic amplifier from the filter pair, employed in the experiment described, was associated with too short a time-constant.

The range of suitable time-constants is, however, very wide. Those used with the long-tailed pair were chosen within this range to give a fairly rapid response time of the servo mechanism and also to permit efficient use of the error signal.

Oscillograms of the growth and decay of the error voltage between the anodes of the long-tailed pair are shown in Figs. 12 and 13 for Gaussian spectra with mean frequencies of 1 and 4kc/s respectively. The inputs to the filter pair were switched to give bursts of 0·15sec duration at balance and at preset errors of 200 and 500c/s. The magnetic-amplifier output was disconnected

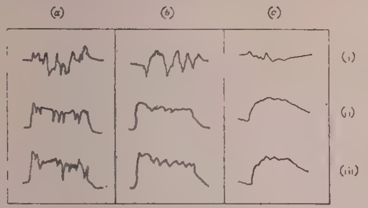


Fig. 12.—Rectifier response to 1 kc/s Gaussian spectrum.

Length of burst: 0·15 sec.
(a) Decay time-constant: 6 millisec.
(b) Decay time-constant: 20 millisec.
(c) Decay time-constant: 100 millisec.
(i) Balance.
(ii) Error: 200 c/s.
(iii) Error: 500 c/s.

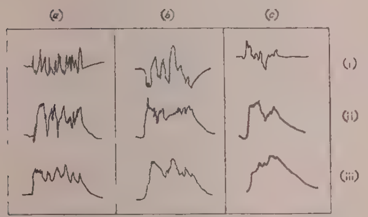


Fig. 13.—Rectifier response to 4 kc/s Gaussian spectrum.

Length of burst: 0·15 sec.
(a) Decay time-constant: 6 millisec.
(b) Decay time-constant: 20 millisec.
(c) Decay time-constant: 100 millisec.
(i) Balance.
(ii) Error: 200 c/s.
(iii) Error: 800 c/s.

during these tests. The conditions shown are for decay time-constants of 6, 20 and 100 millisec, and the growth time-constant in each case was about one-tenth of these values. Longer time-constants may be seen to reduce the ripple on the error voltage at the same time as retarding the response. The magnetic amplifier, which was operated at a power frequency of 400 c/s, had a time-constant of about 10 millisec when a 50-kilohm forcing resistor was connected in series with the control winding. With the shortest time-constant used, the magnetic amplifier would therefore contribute appreciably to the overall response time.

(7) ASYMMETRICAL SIGNAL OR NOISE

The assumed symmetry of spectrum and level-noise characteristic may be inapplicable in certain problems, and in some cases wide filters become inaccurate. The symmetrical Gaussian spectrum and level noise represent the ideal case in which the characteristic frequencies of the power spectrum all coincide. These characteristic frequencies are as follows:

True mean frequency $\dfrac{\int_0^\infty f dW_s(f)}{\int_0^\infty dW_s(f)}$

Modal frequency Frequency at which W_s is a maximum

Median frequency f_{med}, where $\int_0^{f_{med}} dW_s(f) = \int_{f_{med}}^\infty dW_s(f)$

The equipment measures a mean frequency defined by

$$\int_{f_m-f_h}^{f_m}[dW_s(f) + dW_n(f)] = \int_{f_m}^{f_m+f_h}[dW_s(f) + dW_n(f)] \quad . \quad (7)$$

which reduces to an approximation to median frequency in the level-noise case. It may be seen clearly from eqn. (7) that the frequency measured is likely to be neither true mean frequency, nor modal frequency, nor median frequency in the cases of non-level-noise or unsymmetrical signal energy distribution.

Where the noise characteristic is not reasonably level, high noise levels will give unequal signals to two filters of equal response. It is likely, however, that the nature of the noise characteristic will be known and that the maximum tolerable filter bandwidth may be computed for a given maximum tolerable error in measurement and a given maximum noise level.

With noise whose level varies inversely with the frequency, it is clear that the worst error occurs for the lowest mean frequency spectrum. Eqns. (5) and (6) become

$$V_1 \propto \sqrt{\int_{f_m-f_h}^{f_m}\frac{c}{f}df + \int_{f_m-f_h}^{f_m}K\exp\left[-a^2(f_0-f)^2\right]df} \quad . \quad (8)$$

$$V_2 \propto \sqrt{\int_{f_m}^{f_m+f_h}\frac{c}{f}df + \int_{f_m}^{f_m+f_h}K\exp\left[-a^2(f_0-f)^2\right]df} \quad . \quad (9)$$

where K is defined by the signal energy

$$W_s = \int_{-\infty}^{\infty}K\exp\left[-a^2(f_0-f)^2\right]df \qquad \text{as before}$$

and $$W_n = \int_{f_1}^{f_2}\frac{c}{f}df \quad \text{defines } c$$

The extreme frequencies of the pass band are f_1 and f_2. Note that the integral for W_s between f_1 and f_2 will not be appreciably different from that between $-\infty$ and ∞.

The same constant of proportionality applies in eqns. (8) and (9) for any given values of c and K. Hence, for a permitted $x\%$ error due to noise,

$$f_m = (1 + x/100)f_0$$

is the condition in which no error voltage shall appear. If curves of the difference between the exponential integrals varying with f_h are plotted for chosen values of x, and also curves of the difference between the inverse frequency integrals for the chosen worst signal level, they cross one another at the maximum filter bandwidth permitted. Fig. 14 shows the variation of

$$\left[\int_{f_m-f_h}^{f_m}K\exp\left[-a^2(f_0-f)^2\right]df - \int_{f_m}^{f_m+f_h}K\exp\left[-a^2(f_0-f)^2\right]df\right]$$

with f_h for $f_m = 1{\cdot}01 f_0$ and also the variation of

$$\left(\int_{f_m}^{f_m+f_h}\frac{c}{f}df - \int_{f_m+f_h}^{f_m}\frac{c}{f}df\right)$$

with f_h for several signal/noise ratios and $f_1 = 100$ c/s and $f_2 = 8\,000$ c/s. The very surprising result is that either a relatively wide-band filter is permitted, or no solution is possible, depending on the signal/noise ratio. Curves plotted at different errors would reveal the complete picture, but Fig. 14 may easily be interpreted to show that the net error signal of correct sign is a

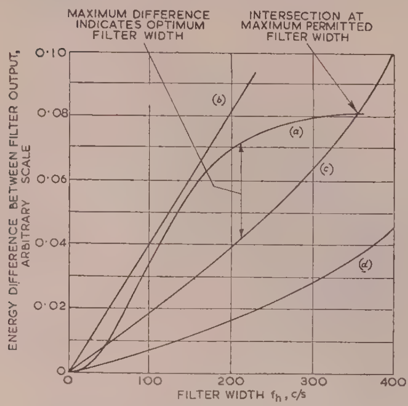


Fig. 14.—Dependence of spectrum (*a*), and noise (*b*), (*c*), (*d*), differential energies on filter width.
(*b*) $W_n = 10\,W_s$.
(*c*) $W_n = 5\,W_s$.
(*d*) $W_n = 2\,W_s$.

maximum where the slopes of the two curves are the same. Curve (*b*) never cuts the spectrum differential curve (*a*), and there is no solution for this noise level with the 1% error chosen. The noise level in curve (*d*) will clearly permit much wider-band filters than curve (*c*), but the optima for the two cases are not greatly different. Noise error signal becomes appreciable for very narrow-band filters and grows continuously while the spectrum error signal has an initial slope of zero. This means that the optimum error signals are not obtained even for this kind of noise distribution with very narrow-band filters. The spectrum error signals for smaller values of x will approximate to similar curves with small ordinates, and since the range of reasonable filter bandwidths indicated by Fig. 14 is large, any suitable pair of filters leads to the same overall conclusions, namely with an inverse frequency distribution of noise there is still a disadvantage in using very narrow-band filters, and for a given noise level there is a necessary error in the measurement of f_0 which cannot be reduced by filter design if the filters touch at f_m.

It may also be seen that a gap between filter pass-bands would offer some improvement, but in the type of system described in the paper, the gap would have a different relative bandwidth according to the different mean frequency spectra and would emphasize any possible asymmetry of the distribution by giving different errors at different frequencies according to changes in noise level.

Asymmetry of the spectrum is amenable to correction by calibration of the measuring arrangement described, provided only that the asymmetry is repeatable and the noise has a level characteristic. Where $2f_h$ is greater than the bandwidth of the spectrum (the energy outside this bandwidth being negligible) the median frequency is measured. However much noise energy is present, equal amounts go to each filter from the signal spectrum at the balance, which for a particular asymmetry at any value of f_0 is a sufficient condition for the reference oscillator to be calibrated in terms of the parameter which generates the spectrum. The addition of f_0 to the reference oscillator frequency will not then yield exactly the mid-frequency of the filters.

(8) CONCLUSIONS

The experimental work indicates that, with the proposed types of filter and method of using the error signal, substantial noise levels can be tolerated. The pull-in range, which is greater than might have been expected, does not deteriorate very rapidly until signal/noise ratios of -10dB are reached. Also the sensitivity does not deteriorate very greatly in the presence of noise, and is such that a measuring accuracy of better than 1% may be expected at all values of f_0. Both properties, i.e. sensitivity and pull-in range, are dependent to a considerable extent on the servo loop injection circuit.

The subtraction and addition systems are similar in performance, and there is no conclusive evidence of the superiority of the latter as a system. The small measured superiority of the addition system, as regards pull-in range, must be attributed to the difference in filter bandwidth and not to the method. However, the experiments were not adaptable to detect errors in mean frequency measurement at particulars values of f_0, caused by intermodulation noise. It must be assumed that these errors could prove serious, and hence the addition system, which is inherently free from them, is preferred. No well-defined optimum for filter bandwidth was indicated by the experiments. The system described is a highly-sensitive and potentially-accurate measuring device possessing advantages of simplicity and not requiring anything remarkable in terms of input-signal quality. This latter feature will doubtless appeal most to engineers working in fields requiring this type of measurement.

(9) ACKNOWLEDGMENTS

The authors are indebted to the Defence Research Board of Canada for permission to publish the paper. They are also grateful to Mr. B. A. Walker for his interest and encouragement and to Mr. J. W. Cox for helpful suggestions on the presentation of the material. The assistance of Mr. W. R. Long in setting up equipment for the experimental work has been appreciated.

(10) REFERENCES

(1) MacLucas, J. L., and Raymond, R. C.: "The Power Spectrum of a Narrow Band Noise passed through a Non-linear Impedance Element," *Journal of Applied Physics*, 1951, **22**, p. 1211.
(2) Kansky, K. G.: "Experimental Investigation of the Characteristics of Certain Types of Noise," *Proceedings of the Institute of Radio Engineers*, 1939, **27**, p. 763.
(3) Rice, S. O.: "Mathematical Analysis of Random Noise," *Bell System Technical Journal*, 1944, **23**, p. 282; and 1945, **24**, p. 46.
"Statistical Properties of a Sine Wave plus Random Noise," *ibid*., 1948, **27**, p. 109.
(4) Fano, R. M.: "Signal-to-noise Ratio in Correlation Detectors," Technical Report No. 186, Research Laboratory of Electronics, Massachusetts Institute of Technology, 19th February, 1951.
(5) Tucker, D. G.: "Rectifier Modulators with Frequency-Selective Terminations," *Proceedings I.E.E.*, Paper No. 863 R, September, 1949 (**96**, Part III, p. 422).
(6) Belevitch, V.: "Linear Theory of Bridge and Ring Modulator Circuits," *Electrical Communication*, 1948, **25**, p. 62.

621.317.333.8 : 621.316.7.078 : 621.373.43

Paper No. 1839 M
May 1955

AUTOMATIC CONTROL AND DISPLAY IN IMPULSE TESTING

By R. F. SAXE, Ph.D.

(*The paper was first received* 9*th November*, 1954, *and in revised form* 26*th January*, 1955.)

SUMMARY

Routine impulse testing is usually carried out using an impulse generator and a cathode-ray oscillograph, the breakdown of the object under test being indicated by the sudden collapse of the wave to zero. The method is limited by the difficulty of visual observation of the cathode-ray tube and by the time delay before the oscillogram may be measured.

Many types of test would be facilitated by the employment of devices to display the peak voltage and duration of the impulse wave without delay, so that the limitations of the cathode-ray oscillograph could be overcome. The paper discusses circuits which enable an impulse generator to be triggered automatically at a preset charging voltage and allow the peak voltage and duration of the wave to be displayed on ordinary meters. In addition, the peak voltage of the wave applied to an object under test can be made visible to several observers without delay and an indication of the breakdown of the object under test can be presented. The circuits also permit automatic control of the test sequence to be achieved if desired.

These circuits are not intended necessarily to replace the cathode-ray oscillograph but to augment it, so that routine testing may be facilitated. One of these circuits has been in operation for many months and has been found to be very reliable.

(1) INTRODUCTION

The testing of insulators, transformers and similar equipment with impulse voltages has become standard practice, the test equipment used consisting mainly of an impulse generator and a cathode-ray oscillograph. The impulse generator is usually triggered by an operator when the charging voltage has reached the desired value, which means that an operator is needed to concentrate on this alone. Although in some cases the peak voltage of the output wave may be determined from a knowledge of the charging voltage of the impulse generator, its accurate determination usually involves the use of a cathode-ray oscillograph. The breakdown of the object under test is also usually determined by means of a cathode-ray oscillograph, on the screen of which the voltage wave is seen to collapse to zero. The measurement of a voltage wave by a cathode-ray oscillograph involves the use of photography to record the trace on the screen and the photographic processing involves a considerable wastage of time before the test results may be assessed.

It would therefore appear to be desirable in many cases that the required information should be displayed immediately after the voltage wave has been applied to the object under test and that the operation of the impulse generator should be made as automatic as possible, so that the operator is free to concentrate on the test as a whole. Furthermore, it is sometimes preferable that the test results should be displayed in such a way that several observers may see them.

The paper shows that the required information may be displayed on ordinary meters, despite the very short times (about 10 microsec) involved in impulse testing, and that almost any desired degree of automatic operation may be incorporated.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Dr. Saxe is in the Department of Electrical Engineering, Queen Mary College, University of London.

(2) CONTROL OF THE IMPULSE GENERATOR

In order to produce a wave of approximately the required voltage amplitude, it is necessary to trigger the impulse generator when the voltage to which it is charged has reached a given value. This may be done by setting the spark-gaps of the generator so that at the required charging voltage the generator will not fire except when the triggering voltage is applied. The voltage to which the first stage of the impulse generator is charged is observed by means of a high-value resistor and microammeter connected in series across the first stage. An operator triggers the impulse generator when the charging voltage has reached the desired level. The accuracy with which the amplitude of the output wave may be controlled is dependent on the accuracy of the series resistance and microammeter and is usually about 1 or 2%, which is considered reasonable in impulse testing.

Automatic triggering of the impulse generator may be achieved by the circuit shown in Fig. 1. In Fig. 1(*a*) (positive-polarity generator charging) the charging voltage is applied to the voltage divider formed by the resistors R_1 and R_2. The potential, V_R,

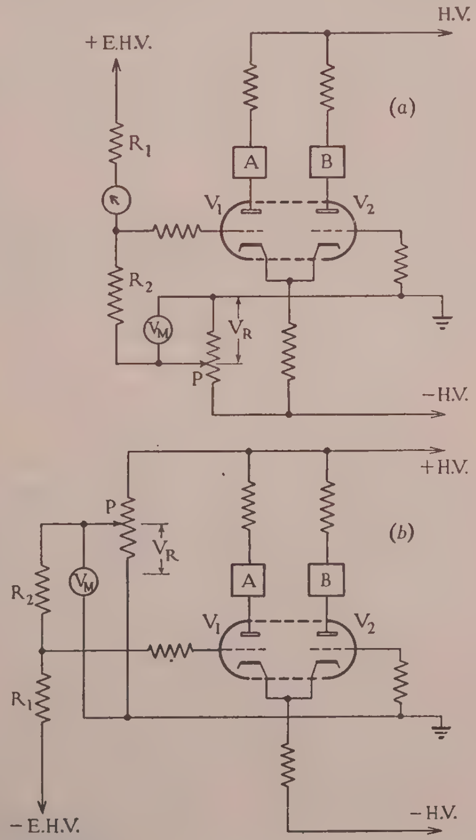


Fig. 1.—Circuits for the automatic triggering of the impulse generator at a preset charging voltage.

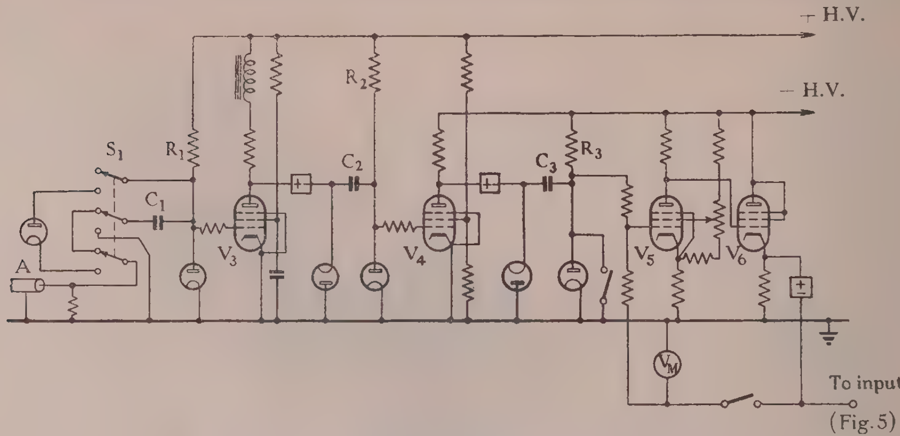


Fig. 2.—Circuit for the display of the peak voltage of an impulse wave.
Switch S_1 is shown in the position for a positive impulse voltage.

of the lower end of R_2 may be adjusted by the potentiometer P and is monitored by the voltmeter V_M. In the waiting condition, valve V_1 is non-conducting and valve V_2 is conducting. When the charging voltage has risen to a value V such that $VR_2/(R_1 + R_2) = V_R$ then V_1 will become conducting and V_2 will become non-conducting. The change-over of current from V_2 to V_1 will occur over a range of grid voltage of about 2 volts if high-slope pentodes are used. This will correspond to a triggering uncertainty of 1% if V_R is 200 volts. In practice it is found that the uncertainty of the triggering point is less than this but that a slight drift of the exact triggering point is apt to occur with ageing of the valves. It is considered, however, that the uncertainty in triggering voltage is within the normal tolerance for impulse testing and that the use of more complicated circuits is not justified.

The connections for the case of negative-polarity generator charging are shown in Fig. 1(*b*).

The contacts on either relay A or relay B are used to apply the triggering pulse to the impulse generator. Extra contacts on either relay may be used to discontinue the charging of the generator when the change-over of current in valves V_1 and V_2 occurs, thereby preventing self-triggering of the generator at a higher voltage if the normal triggering mechanism fails to operate.

(3) DISPLAY OF THE WAVE PARAMETERS

(3.1) Peak Voltage of the Wave

The peak voltage of the impulse wave may be measured by reference to the charging voltage of the impulse generator, by the use of a calibrating sphere-gap or by recording the wave on a cathode-ray oscillograph. The first of these methods is applicable only when the effect of the output circuit on the waveform is known, while the use of a sphere-gap requires the application of a large number of impulses to the object under test, which may be undesirable. The use of a cathode-ray oscillograph enables an accuracy of about 1% to be achieved and with care the result is beyond doubt. However, visual observation of the wave on the screen is difficult and is usually restricted to one observer. The delay before the oscillogram is available for measurement is often undesirable, and it would therefore appear that there is a need for an instrument which would present the peak voltage of the wave immediately after its application. Such a device is shown in Fig. 2.

The voltage wave, which may be of either polarity, is applied after division by a suitable voltage divider to the input plug, A, and charges the capacitor C_1 to a negative voltage equal to th peak voltage of the wave. This negative voltage renders valve V non-conducting until it is removed by leakage through the resisto R_1. Since R_1 is connected to the high-voltage source and th negative voltage impressed on C_1 is arranged to be small com pared with the high voltage, the rate of leakage of voltag from C_1 is sensibly constant. As a result, the time for which V is non-conducting is directly proportional to the peak voltage o the applied wave.

While V_3 is non-conducting, the capacitor C_2 is being charge by current flow through the anode load of V_3. When V_3 agai becomes conducting, the voltage to which C_2 has been charge appears as a negative voltage on the grid of V_4 and leaks awa through R_2. Since the time-constant of the second stage (C_2R_2 is greater than that of the first stage (C_1R_1), a time expansion ha been achieved. A further time expansion is achieved in the thir stage (C_3R_3).

The peak voltage to which C_3 is charged is measured by "see-saw" circuit consisting of valves V_5 and V_6 and associate components, the voltage being read on the voltmeter V_M. Th rate at which this reading decays is so slow that, provided tha the observation is made within a few seconds, the loss o accuracy is negligible.

(3.2) Time Lags*

It will be noticed that the circuit described in the previou Section depends for its action on a direct relationship between

* Moody, N. F.: "Time Expansion for Millimicrosecond Pulse Intervals," *Electron Engineering*, 1952, 24, p. 289.

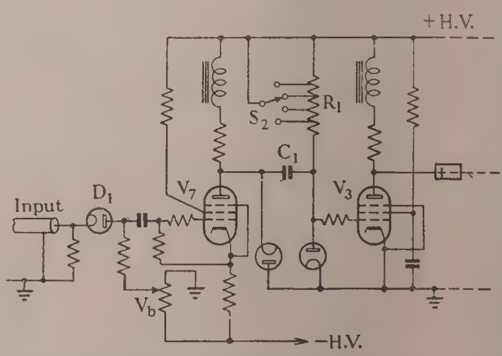


Fig. 3.—Modification of the circuit shown in Fig. 2 for the display c time-lags.

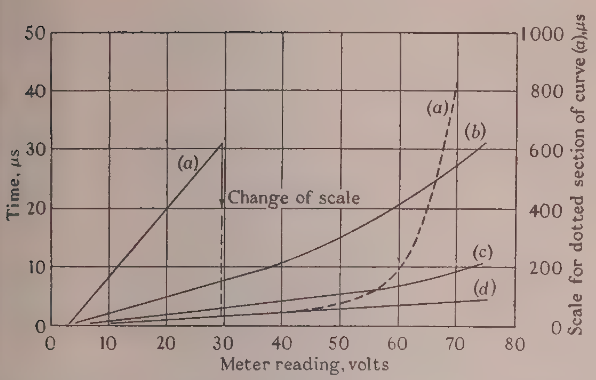


Fig. 4.—Typical calibration curves for the circuit for the display of time-lags.

Curves (*a*), (*b*), (*c*) and (*d*) correspond to different positions of S_2 (Fig. 3).

time interval and a voltage. If, therefore, the circuit shown in Fig. 2 is preceded by a valve connected as shown in Fig. 3, the time duration of a voltage wave may be registered on the voltmeter, V_M. The wave is applied to V_7 through a diode D_1 biased to a voltage V_b, and the time duration which is recorded is that for which the wave exceeds V_b.

(4.1) Repeated Impulses at the Same Voltage

In some applications, e.g. insulator testing, it may be required to apply several waves to the object under test at the same peak voltage. In this case the triggering circuit is preset to trigger the generator at the required charging voltage, and when the charging set is switched on, the generator continues to produce impulses at approximately regular intervals until it is stopped. Since the time-constants of the display circuits are so long, a resetting device such as that shown in Fig. 5 must be provided. The relaxation time of valve V_9 is arranged so that the meters may be read before the resetting actions occur, and the impulse generator must be adjusted so that the impulses do not occur so rapidly that the resetting action is not completed in time.

The inclusion of a relay in the circuit of V_8 enables the charging of the impulse generator to be interrupted whenever a reading appears on the meters and to be recommenced only when the resetting actions are complete. In this way the speed with which the generator is charged is rendered unimportant.

(4.2) One Impulse only at any Voltage Setting

The requirement that the impulse generator should produce one impulse only at a given voltage setting may be met by the circuit shown in Fig. 6. A is a self-held relay fed from the high-voltage source through resistor R and the contacts on relay B in the circuit of V_8. When an impulse occurs relay B is energized,

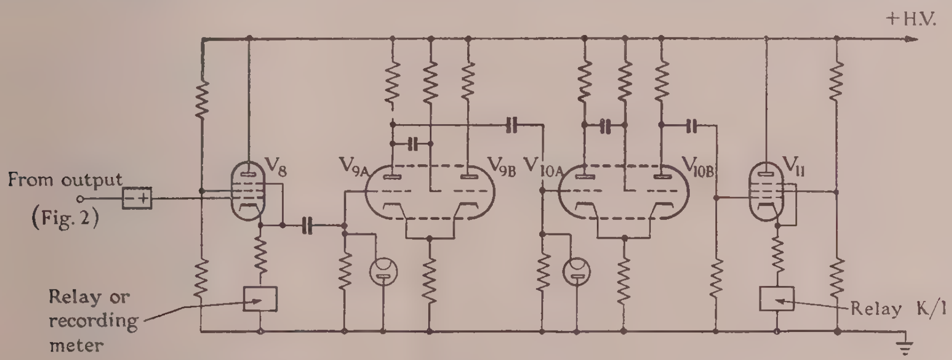


Fig. 5.—Circuit for automatic resetting of the display circuits.

The circuit shown in Fig. 3 will function only when the wave applied to D_1 is of negative polarity. If the impulse generator produces a positive wave, a valve circuit must be included to invert the input to V_5.

In addition, a switch, S_2, is included to vary R_2 (Fig. 2), so that different ranges of time may be covered. Typical calibration curves are shown in Fig. 4. The non-linearity which is apparent on some of the ranges could readily be removed if desired, but is considered to be useful in that a greater range of times may be covered on the one range than would be the case with a linear law.

For routine testing with a 1/50 microsec wave, for example, the range switch and bias voltage V_b would be set so that a full-scale reading corresponding to, say, 100 microsec was obtained in the absence of a breakdown. Under these conditions, when a breakdown occurs the meter will give a reading of considerably less than full scale, since a time lag to breakdown longer than 50 microsec is unlikely.

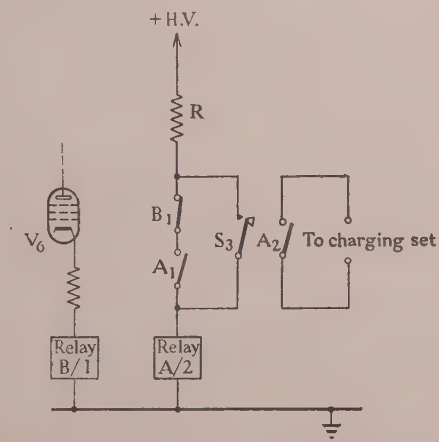


Fig. 6.—Relay connections for producing one impulse only at a given voltage setting.

(4) AUTOMATIC OPERATION

The degree of automatic operation desirable obviously depends on the particular application envisaged; in this Section, therefore, various possibilities to illustrate the general principles will be presented.

and by breaking the circuit causes relay A to open, thereby interrupting the charging of the generator. The charging may be recommenced by closing the switch S_3, and this may also be made to reset the display circuits.

(4.3) Repeated Impulses at the Same Voltage until a Breakdown Occurs

It may be required to apply impulses at a given voltage to the object under test until a breakdown occurs and then to discontinue the test. This may be achieved by a circuit (Fig. 7) similar to that shown in Fig. 5, except that an extra diode, D_2, is inserted. When no breakdown occurs, a full-scale deflection is obtained on the time-lag meter, and the voltage produced across R_1 is sufficient to overcome the bias on D_2 and to cause the resetting action to occur. The charging is interrupted by the opening of a relay, as described above, and is recommenced only when the resetting action is complete.

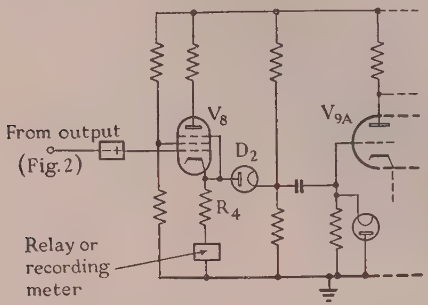


Fig. 7.—Circuit modification to allow repeated impulses to be discontinued if and when breakdown occurs.

When breakdown occurs the voltage appearing across R_4 is considerably smaller than in the absence of breakdown, and may be arranged to be insufficient to overcome the bias on D_2. As a result, the charging of the generator is discontinued and the resetting action does not occur.

(4.4) Other Forms of Automatic Operation

Other forms of automatic operation will be obvious. For example, the use of a uniselector would enable the triggering level to be increased with each impulse, so that impulses of increasing peak voltage would be applied to the object under test. The use of a counter or uniselector would enable a given number of impulses of a given voltage to be applied before the test was automatically stopped or before the applied voltage was increased.

(5) ACKNOWLEDGMENTS

The author wishes to acknowledge the help given by Professor W. J. John, in whose laboratories the work was performed. The work was carried out during the tenure of an I.C.I. Research Fellowship, which is also gratefully acknowledged.

621.376.54

Paper No. 1812 R
May 1955

A 24-CHANNEL PULSE-TIME MODULATION SYSTEM

By R. F. B. SPEED, B.A.

(*The paper was first received* 2*nd October, and in revised form* 11*th December*, 1954.)

SUMMARY

In this paper a communication system is described which uses pulse-time modulation. It provides 24 channels of the quality recommended by the C.C.I.F. for the transmission of speech over international trunk circuits.

The pulse-multiplex terminals operate in conjunction with radio equipment operating in the 2000 Mc/s band; this is not described here.

The multiplexing equipment and its mode of operation are described in some detail—in particular the distributor and modulator.

Performance figures are given.

(1) INTRODUCTION

The idea behind the use of time division for the multiplexing of signals is now quite old, having been used as long ago as 1874 by Baudot for telegraphy. In 1919 H. J. Round took out a patent on pulse-width modulation,[1] although he does not seem to have envisaged the interlacing of different channels. However, in 1927 J. M. Loeb produced a system employing interlaced amplitude-modulated pulses.[2] In 1935 R. D. Kell took out a patent on the use of time and width modulation,[3] and in 1937 E. M. Deloraine and A. H. Reeves patented a system using interlaced time-modulated pulses;[4] this seems to have been the first practical system using pulse-time modulation.

Since 1937 a number of systems have been produced using time or width modulation, the general tendency being towards an improved performance and a reduction in the number of valves required (e.g. References 5, 6 and 7).

The pulse system to be described here, operating in conjunction with a radio system, provides the quality of transmission associated with long coaxial cables, and is simple both in its construction and in operation.

(2) GENERAL DESCRIPTION

Although only the pulse modulation, demodulation and multiplexing equipment is to be described in the paper, a brief description of the whole system is given here so that the overall performance may be understood and the relationship between the pulse and radio equipment made clear.

One complete system provides 24 one-way channels of the quality required by the C.C.I.F. for transmission of speech over long coaxial cables.[8]

The main problem involved in long-distance communication is that of achieving the required signal/noise ratio. The noise in the received signal is due to a variety of causes which may conveniently be divided into two parts—the noise caused in transmission by the radio transmitter and receiver, and that introduced by the terminal equipment, in this case the pulse modulators and demodulators.

The radio link uses frequency modulation and its characteristics may be briefly summarized as follows:

Frequencies	in the 2 000 Mc/s band
Peak-to-peak deviation	12 Mc/s
Transmitted power	1 watt
Aerial gain: transmitter and receiver (each) ..	30 dB
Receiver noise factor..	14 dB
Average repeater spacing	30 miles

Making due allowance for losses in feeders, etc., and for fading this gives, under the worst conditions, a carrier/noise ratio at the receiver of 23 dB. As shown in the Appendix, if such a radio link is fully modulated by a 24-channel system of the type to be described, the signal/noise ratio in the demodulated channels will be 40·8 dB better than the radio carrier/noise ratio by virtue of the combined pulse-time modulation and frequency-modulation improvements. Hence, under the worst conditions, the signal/noise ratio will be approximately 64 dB. In practice, when a large number of repeaters are used, the probability of serious fading occurring on many of the sections at one time is very low: when due allowance is made for this it is found that the C.C.I.F. requirements for noise power in the output channels are met over a 2 500 km circuit.

The number of channels is fixed at 24 so that the system should be equivalent to two 12-channel groups.

In a pulse system the repetition frequency must be at least twice that of the highest frequency to be transmitted: in this case the band to be used lies between 300 c/s and 3·4 kc/s, so the repetition frequency must be above 6·8 kc/s. The frequency of 8 kc/s was chosen for a variety of reasons, the most important being that after the demodulation process it is necessary to attenuate all frequencies above half the repetition frequency, while on the other hand no attenuation must be allowed to take place in the pass band; if the highest frequency to be transmitted and half the repetition frequency are suitably spaced the desired result can be achieved with a simple filter.

In addition to the modulated pulses a synchronizing pulse must be provided so that the receiver oscillator can be locked to the incoming signal in the correct phase. The use of pulses of a different shape for this purpose is the obvious solution; this has the serious disadvantage, however, that in any reshaping of the pulses which may be necessary, at repeaters for example, the synchronizing pulses would have to be treated differently. Instead, a double pulse is used; two pulses of the same shape as the channel pulses are separated by about one pulse length and are identified at the receiver by this property. An additional channel interval is provided for the synchronizing pulse, bringing the number to 25. Since an odd number of channels is inconvenient for timing purposes, a spare channel interval is provided, making the total 26. This is normally left vacant but could be used for a pulse carrying an engineers' channel.

The length of pulse used is affected by a number of factors. If the pulse is wide the possible time deviation in the channel interval is reduced; if it is made narrower the video bandwidth required is increased thus increasing the noise, this situation being worsened by the fact that frequency modulation with its well-known triangular noise spectrum[9] is being used. As a compromise a pulse length of 0·5 microsec was chosen, this involves a video bandwidth of about 2·0 Mc/s. Cross-talk considerations also preclude the use of a broad pulse with a long "tail."

If the pulses are allowed to be modulated over the entire channel interval severe crosstalk may be caused in adjacent channels if a pulse comes very near to the boundary between them. Experience shows that, if the excursions of the pulses are limited to 70% of the total channel width, no crosstalk is caused by this effect. This, incidentally, also ensures that two pulses

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
The paper is a communication from the Staff of the Research Laboratories of The General Electric Company, Limited, Wembley, England.

from adjacent channels can never come close enough to one another to simulate a synchronizing pulse.

Various important features of the pulse equipment may conveniently be mentioned here before proceeding to a more detailed description.

One of the most important components in a pulse system is the distributor by which the timing of the channel pulses is controlled. Various types have been used in the past, some active, such as cathode-ray tubes and chains of valves, and some passive, such as delay lines. Both types have their disadvantages; the active type may be dependent upon special cathode-ray tubes or be such that, if one valve ceases to function, the whole chain subsequent to it ceases to work. Delay lines of the lumped-constant type normally used are never perfect, so the output pulses almost invariably have to be reshaped in some way requiring the use of extra valves. In the present system a passive network is used to phase-shift a sine wave at the repetition frequency (8 kc/s). The outputs from this network are always pure sine waves, and by using close-tolerance components and high-Q-factor coils they are all of very nearly the same amplitude. The nearly linear section of the sine wave about the point of steepest slope is used directly in the modulation process, as explained later, the time modulation of the pulses depending directly on the slope of the sine wave at this point. This slope depends directly on the amplitude of the sine wave, which is kept to within 0·5 dB of its input value. The effect of the small variation can be compensated for in the adjustable output amplifiers which follow the demodulators.

The same type of distributor is used in the receiver, although in this case not so directly, since a gating pulse has to be developed from each output. However, from a purely practical point of view it is still very convenient, since only one section of the network has to be used for each tapping, as opposed to a minimum of three or four in a delay line for delaying pulses.

A channel of greater bandwidth (7·4 kc/s) suitable for medium-quality music transmission may be provided by this equipment by modulating two equally spaced channels (say 1 and 14) with the same signal; this effectually doubles the repetition frequency and so enables the greater bandwidth to be transmitted.

The total number of valves used in the pulse equipment (one transmitter and one receiver) is 106, corresponding to just under 4·5 valves per one-way channel.

(3) THE TRANSMITTER

The transmitter of a pulse system has to perform the following functions:

(*a*) To modulate each of a number of pulse trains (all of the same repetition frequency) with an audio signal which it is desired to transmit.

(*b*) To interlace these pulse trains so that they do not interfere with one another.

(*c*) To add some kind of synchronizing signal so that the receiver is able to sort out the incoming signals.

The layout used in the equipment to be described is shown in block form in Fig. 1.

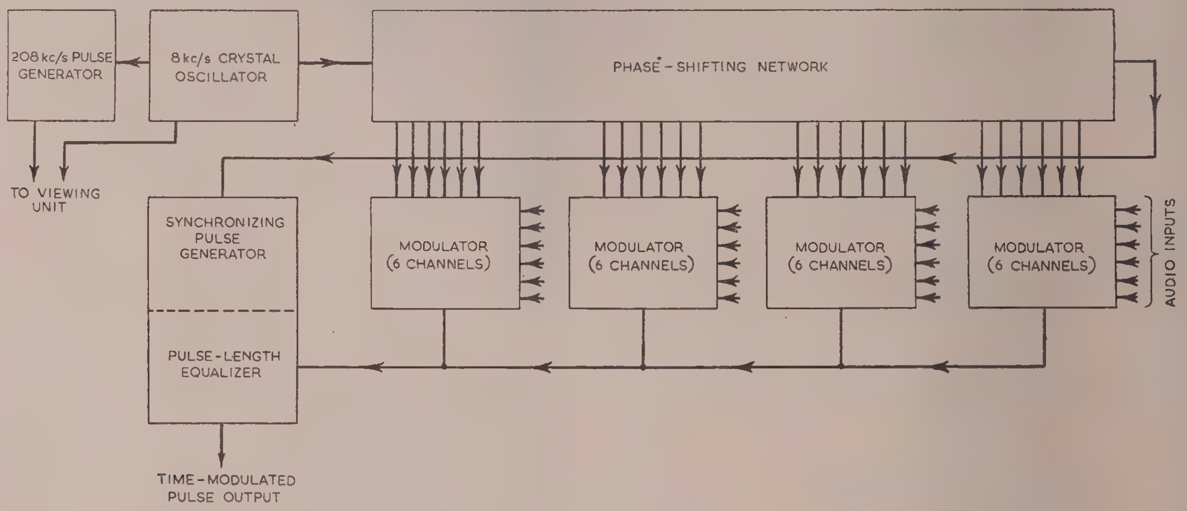


Fig. 1.—The transmitter.

The basic frequency is controlled by the 8 kc/s crystal-controlled oscillator. This provides three outputs: one to a pulse generator which gives an output of pulses at 208 kc/s (i.e. 26 × 8 kc/s) for use as timing calibration on a display unit. The second output is used to synchronize the time-base of this display unit. The third, of carefully controlled amplitude, is fed into a phase-shift network. This network is arranged so that the phase shift between successive tapping is $2\pi/26$ rad at 8 kc/s. The attenuation of the network is kept very low.

Successive tappings from this network are connected to channel modulators which are arranged in batches of six for convenience in construction.

From the end of the phase-shift network an output is taken to the synchronizing-pulse generator, where are generated the double synchronizing pulses previously mentioned.

The pulses from the four 6-channel modulators together with the synchronizing pulses are interlaced and shaped in a pulse-length equalizer, so as to remove any irregularities in amplitude and duration due to component variations, etc. The resultant pulses are then passed on to the radio equipment.

Setting up this equipment, in so far as the timing of the waveforms is concerned, is very simply achieved by using a display unit whose time-base is locked to an output from the master oscillator as previously mentioned. The 208 kc/s pulses are used to mark the channel boundaries and enable all the timing to be adjusted to the required degree of accuracy.

(3.1) The Master Oscillator

The oscillator, which runs at 8 kc/s, is crystal controlled and of conventional design. The use of a crystal ensures that good fre-

quency stability is obtained without the use of stabilized power supplies and temperature-compensated tuned circuits. Calculation showed that under extreme conditions the synchronization circuit at a receiver terminal, possibly hundreds of miles away, would not be able to keep the receiver oscillator accurately phased if the transmitter frequency were not carefully controlled.

The output to the phase-shift network is controlled at 30 volts peak amplitude, a gas-discharge stabilizer valve being used to provide an accurate reference voltage independent of power-supply variations.

(3.2) The Distributor

The phase-shift network into which the sine wave is fed consists of cascaded π-sections of constant-k low- (or high-) pass filters

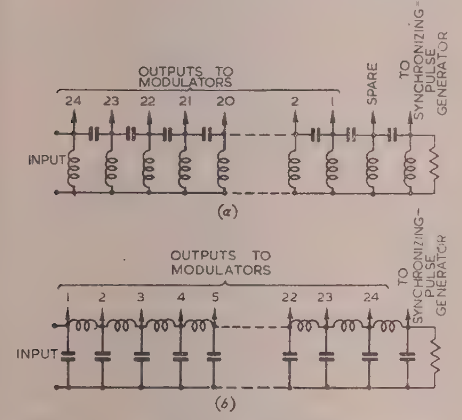


Fig. 2.—Transmitter phase-shifting networks.
(*a*) Phase-advancing network.
(*b*) Phase-retarding network.

(Fig. 2). Each section gives a phase shift corresponding to $2\pi/26$ rad at 8 kc/s, i.e. a phase shift corresponding to one channel interval. There are 25 sections, so that all possible phases may be obtained. The network is accurately terminated in its characteristic impedance at 8 kc/s. By the use of iron-dust cores for the inductances, high Q-factors are obtained and the loss in the network is kept to about 0·5 dB overall.

Both low- and high-pass networks have been used with equally good results, but the low-pass type is preferred as having more convenient component values; also, since it is the same as a conventional delay line, it is easy to check that the termination is approximately correct by passing a low-frequency square wave down it, and examining the input for reflections.

(3.3) The Modulator

A method which is frequently used for generating time-modulated pulses relies on the principle of comparing the signal with which the pulses are to be modulated with a sawtooth waveform which is repeated with the desired pulse-repetition frequency. In this equipment a small section of the 8 kc/s sine wave about the cross-over (point of maximum slope) is used. Near the cross-over a sine wave approximates to a straight line, and elementary calculation shows that the non-linearity produces no second-harmonic distortion, and that the third-harmonic distortion is 70 dB below the fundamental with the maximum modulation. This amount of distortion is less than is likely to be produced by non-linearity in a sawtooth.

The modulator circuit is shown in Fig. 3. Valves V1(*a*) and V1(*b*) constitute the modulator and are triodes connected with a common cathode load and feedback between one anode and the other control grid. The audio signal with which the pulse train is to be modulated is applied to the grid of V1(*a*). The audio signal is limited in amplitude by the germanium diodes G1 and G2 in order to prevent over-modulation of the pulse train. To the grid of V1(*b*) is applied the 8 kc/s sine wave from the appropriate tapping of the phase-shift network [see Fig. 4(*b*)]. It was found to be advisable to feed this sine wave via the cathode follower V2 in order to eliminate a tendency for waveforms to feed back into the phase-shift network from the modulator and cause crosstalk into the other channels.

The operation of the modulator may best be described as follows. Consider the time when V1(*a*) is conducting and the sine wave on the grid of V1(*b*) is rising from its most negative value [Fig. 4(*b*)]; as it approaches to within a few volts of the potential on the grid of V1(*a*), which is determined by the incoming audio signal, V1(*b*) will start to conduct, and its anode potential will fall; this fall will be communicated to the grid of V1(*a*) via the condenser C1, thus accelerating the transfer of current from V1(*a*) to V1(*b*). This change-over is very rapid (less than 0·1 microsec). As the voltage on the grid of V1(*b*) falls again, after the peak of the sine wave, current is again transferred to V1(*a*).

The output is taken from the anode of V1(*a*) and differentiated by the inductor L1, in parallel with which is the germanium diode, G3, so arranged that the only pulses produced across L1 are positive-going and will be derived from the rising edge of the rectangular output waveform. The inductor L1 is chosen

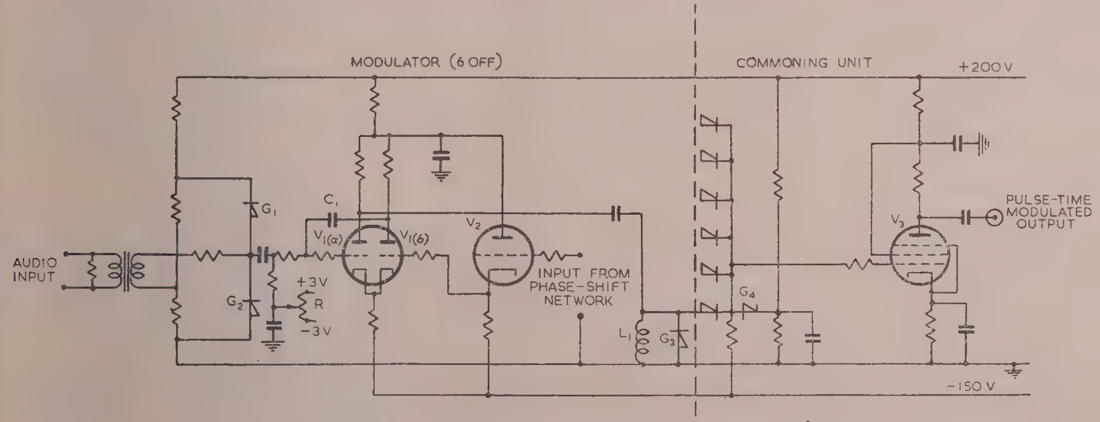


Fig. 3.—The modulator.

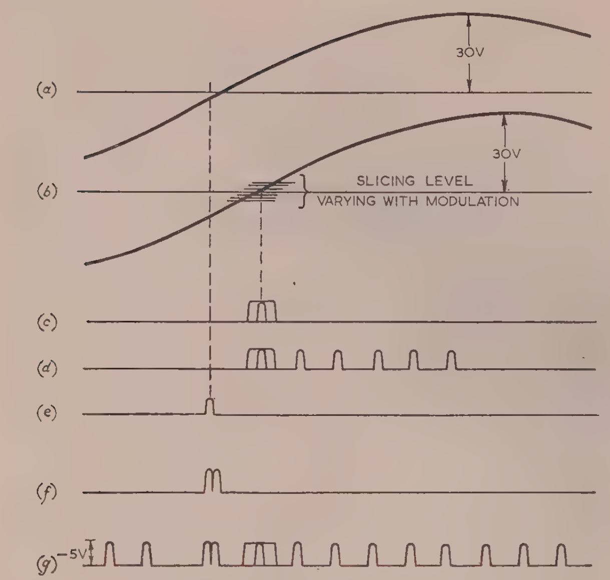


Fig. 4.—Transmitter waveforms.

(a) Oscillator output and phased sine wave for synchronizing pulse.
(b) Phased sine wave for channel 1.
(c) Channel 1.
(d) Channels 1, 2, 3, 4, 5, 6.
(e) Trigger for synchronizing pulse.
(f) Synchronizing pulse.
(g) Output pulses.

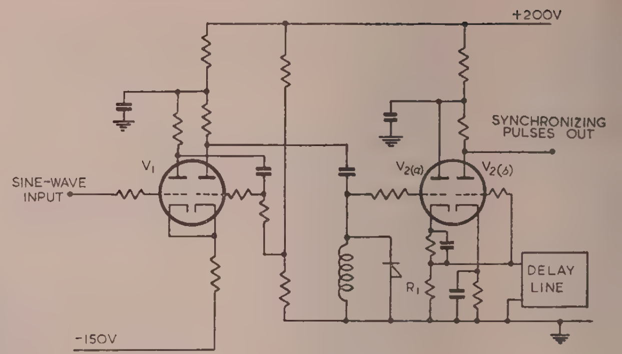


Fig. 5.—Synchronizing pulse-generator.

so that—together with the stray capacitances—it will resonate to give 0·5 microsec pulses [Fig. 4(c)].

The pulses from six of these modulators are combined via germanium diodes which are used to isolate the ringing circuits from one another [Fig. 4(d)].

The diode G4 is so biased that it slices the pulses about one volt above earth potential and so removes any overshoots or stray waveforms which might introduce crosstalk.

The amplifier valve V3 provides a low-impedance output from which the pulses are taken to the final unit, where they are combined with the pulses from the other three 6-channel modulators and the synchronizing pulses. The timing of the output pulses from each channel modulator is controlled by the phase of the sine wave which is fed to it. In addition, a fine adjustment is provided by the potentiometer R1, which varies the mean potential on the grid of V1(a) and so permits a small variation in the section of the sine wave used.

(3.4) Synchronizing-Pulse Generator and Output Stage

With reference to Fig. 5, a 0·5 microsec pulse [Fig. 4(e)] is generated by valve V1 and its associated circuits in exactly the same way as in a modulator. No adjustment is provided for this pulse, since it is used as the reference to which the timing of all the channel pulses is adjusted; it is fed via the cathode-follower V2(a) to the delay line. The delay line is open-circuited at its far end, so that the input pulses are reflected with the same polarity and return to the input, which is terminated by R1, after one microsecond. The resultant double pulse is cleaned up by the Class C amplifier V2(b) [Fig. 4(f)].

The synchronizing pulses are then combined with the channel pulses and applied to a length-equalizing circuit consisting of a monostable multivibrator triggered by the leading edges of the pulses, whose recovery time is adjusted to be 0·5 microsec. In the output from this multivibrator is a pulse transformer; this is used so that an ordinary low-power valve can provide 5-volt pulses in an impedance of 75 ohms which can be fed via a coaxial cable to the radio apparatus.

(4) THE RECEIVER

The receiver must perform three functions: (a) it must select the synchronizing pulse from the incoming pulse train and lock the waveforms generated to it; (b) it must gate out the separate channel pulse-trains; and (c) it must demodulate each pulse train and amplify the resulting audio signal.

From Fig. 6 it will be seen that the pulse train is fed into the synchronizing-pulse separator; in this unit the synchronizing pulse is isolated from the pulse train and compared in a phase-sensitive device with a sawtooth waveform generated from the oscillator sine wave. This in turn generates a bias voltage which is applied to a reactance valve which controls the oscillator frequency.

A further output from the oscillator goes to a pulse generator: in this are generated pulses at a repetition frequency of 208 kc/s. These are so timed as to appear at the channel boundaries.

The phase-shift network is the same as in the transmitter, except that to each output are added the 208 kc/s pulses. These outputs are used to generate gating pulses for each channel, the accurate timing of these being achieved by the superimposed 208 kc/s pulses. The time-modulated pulses are then gated by these pulses in a germanium-diode gate and applied to the demodulator valve. The demodulator converts the time-modulated pulses into width-modulated pulses, which are passed through a low-pass filter in order to reproduce the original signal and finally amplified to give the required output level.

(4.1) Synchronization

Demodulation of the time-modulated pulses is achieved by converting these into width-modulated pulses. The timing of their leading edges is determined by the time-modulated pulses, and that of the trailing edges by waveforms generated in the receiver. These width-modulated pulses are then demodulated by being passed through a low-pass filter. Clearly any noise that appears on the trailing edges of these will appear in this output. The synchronizing pulses will be distorted by random noise in the same way as the received time-modulated pulses; if, therefore, the receiver waveforms are directly locked to the synchronizing pulses, there will be as much disturbance due to random noise on the trailing edges of the width-modulated pulses as on the leading edges, i.e. the level of noise in the output will be 3 dB worse than if the trailing edges were noise free. This figure of 3 dB assumes that there is no correlation between the noise on the two edges. It is therefore desirable to use a method of synchronization whose operation is somewhat analogous to the

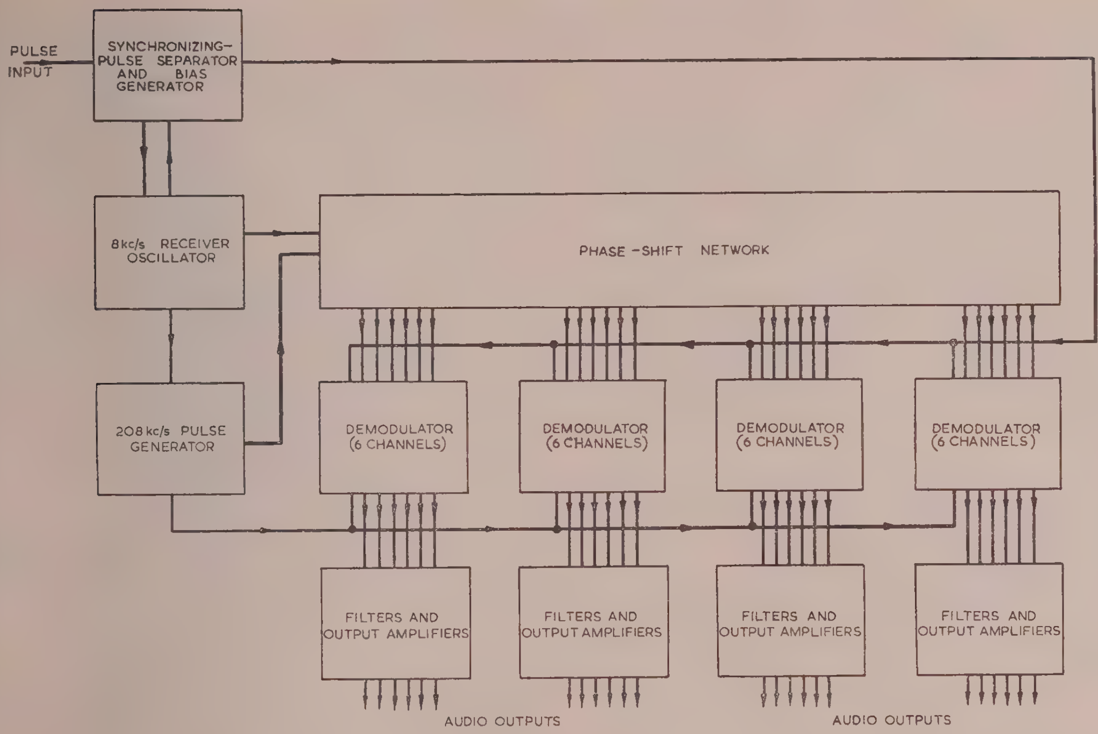


Fig. 6.—The receiver.

"flywheel synchronization"[10] used in some television-receiver circuits.

In Fig. 7, the synchronizing pulse is separated from the pulse train by applying the latter to the delay line, open-circuited at the far end. The delay is such that the reflected first synchronizing pulse coincides with the second one, thus producing a double-height pulse which is separated from the rest of the pulse train in the Class C amplifier V1. In the anode of this potential of the sawtooth at the instant of the pulse. When there is no pulse present the diodes will be biased off and the charge on the condenser C1 will remain unchanged until the next pulse. As the phase of the 8 kc/s oscillator varies in relation to that of the incoming synchronizing pulse the voltage on the condenser C1 will vary. The voltage on C1 is fed via R_1 to the grid of the reactance valve V2. Since the time-constant R_1C_2 is very large, variations due to noise on the synchronizing pulses

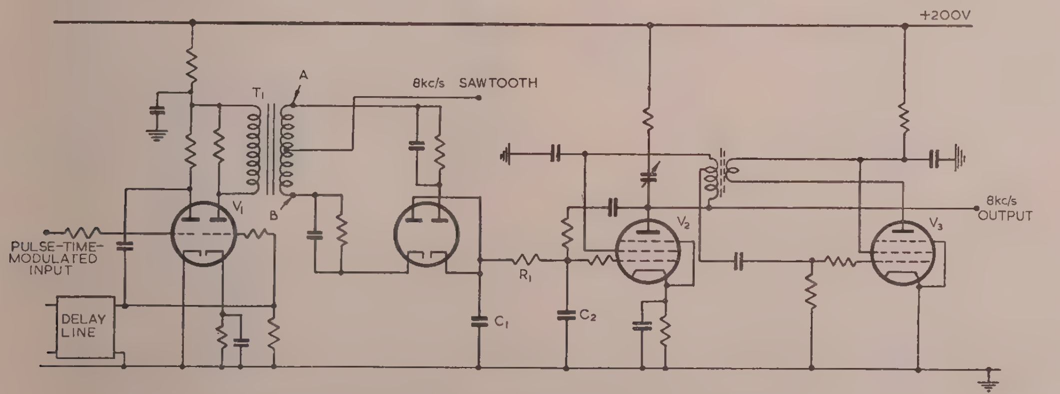


Fig. 7.—Synchronizing pulse separator and locked oscillator.

valve is the diode-clamp transformer T1. To the centre tapping of this transformer is fed a sawtooth waveform generated from the 8 kc/s sine wave; this is arranged to have a flyback of about 5 microsec. The waveform at A is shown in Fig. 8(*b*), the waveform at B is identical except for the polarity of the superimposed pulse. During the presence of this pulse, current will flow around the clamp circuit and the condenser C1 will charge to the will be effectively suppressed. The reactance valve V2 controls the frequency of the *LC* oscillator and enables an accurately phased lock to be achieved for periods of thousands of hours.

(4.2) The Receiver Distributor

The receiver distributor is exactly the same as in the transmitter except that a train of negative pulses, with a repetition frequency

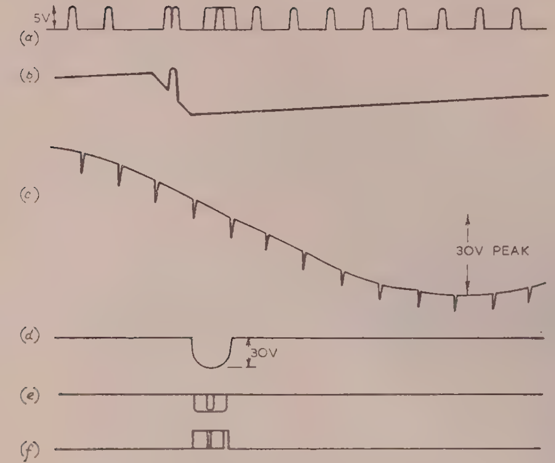


Fig. 8.—Receiver waveforms.

(*a*) Input pulses.
(*b*) Clamping waveform (sawtooth + pulse).
(*c*) Channel 1 selector sign wave with pulses.
(*d*) Channel-selecting pulse.
(*e*) Selector channel 1.
(*f*) Pulse-width modulation on grid of output multivibrator.

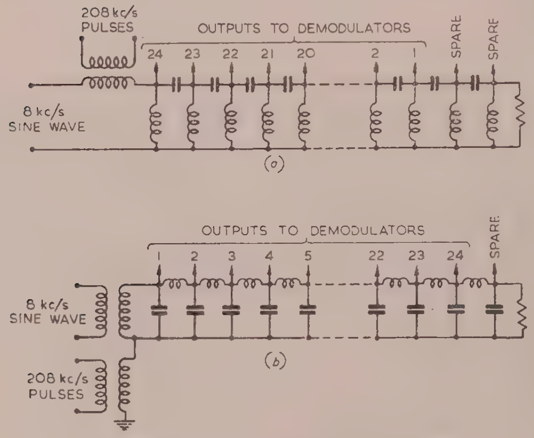


Fig. 9.—Receiver phase-shifting networks.

(*a*) Phase advancing network.
(*b*) Phase retarding network.

of 208 kc/s and fixed phasing, is superimposed on the outputs. The most convenient way of achieving this is shown in Fig. 9. In the high-pass network the pulses suffer negligible delay, since their repetition frequency is very far removed from the cut-off frequency of the network (about 900 c/s). In the low-pass network this form of connection cannot be used since the pulses would be severely attenuated owing to its very low cut-off frequency, so the arrangement of Fig. 9(*b*) is used.

(4.3) The Demodulator

The demodulator may conveniently be divided into three stages: a pulse generator controlled by the sine wave from the phase-shift network; a gate which is operated by the pulse from it, and which lets through the pulses belonging to the channel to be demodulated; and the actual demodulator which converts the time-modulated pulses into width-modulated pulses prior to passing them through a low-pass filter, the output of which is the required audio signal.

The circuit used is shown in Fig. 10. Valves V1(*a*) and V1(*b*) are triodes connected in a similar fashion to those used in the modulators. To this circuit are fed the sine wave and pulses from the phase-shift network [Fig. 8(*c*)]. Consider the time when the voltage on the grid of V1(*a*) is falling: the first negative pulse, which lowers the potential on this grid to within a few volts of that on the grid of V1(*b*), will allow V1(*b*) to start to conduct. Owing to the feedback action, current will be rapidly transferred to V1(*b*), and this will remain conducting until such time as the voltage on the grid of V1(*a*) rises sufficiently to reverse the process.

It will be noted that the time when V1(*b*) starts to conduct is accurately determined by the negative pulses which are themselves adjusted to occur at the channel boundaries. The output from the anode of V1(*b*) is differentiated by the tuned circuit L_1C_1, across which is the germanium diode G1, so arranged that the negative pulse of Fig. 8(*d*) appears on the output. The pulses are of 4·81 microsec duration.[13]

These large negative pulses are applied to a germanium-diode gate. Negative time-modulated pulses are fed to this as shown; so long as the germanium diode G2 is conducting these pulses will merely switch off the germanium diode G3, and the voltage at point A will remain practically constant. However, when the negative-gating pulse appears on G2 the voltage at point A will follow the modulated pulse which arrives during the interval when G2 is switched off.

The time-modulated pulses which are let through by the gating pulses are applied to the grid of V2(*a*). The valves V2(*a*) and V2(*b*) are again connected in a similar fashion to the modulator, and their grids are connected to stabilized voltages as shown. Negative pulses at 208 kc/s are applied via the germanium diode

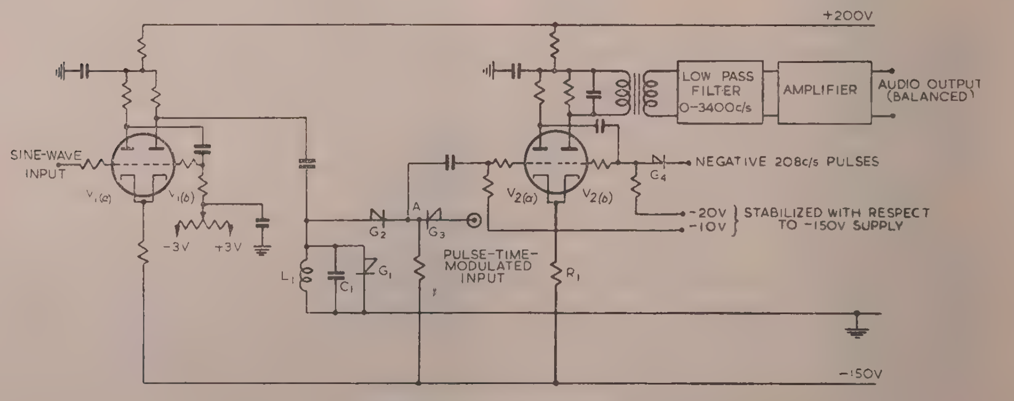


Fig. 10.—The demodulator.

G4 to the grid of V2(b). Between these pulses the grid of V2(b) is held at a potential which is stabilized 150 volts positive to the negative supply, i.e. at approximately earth potential.

Normally, valve V2(a) is conducting and V2(b) is cut off. On the arrival of a negative gated modulated pulse on the grid of V2(a) the current through this valve falls and the potential on its anode rises; the potential on the grid of V2(b) rises with it until G4 starts to conduct and holds it at about earth potential. Owing to the feedback action, V2(a) is now cut off and the current through V2(b) is entirely determined by R1 and the stabilized potential on its grid. On the arrival of a negative pulse from the 208 kc/s pulse train on the germanium diode G4, this diode is cut off and the potential on the grid of V2(b) falls so that V2(a) starts to conduct again, its anode potential rises, driving down the potential on the grid of V2(b) still further, and thus the original conditions are restored.

The current pulse, which has been passed through V2(b), had its leading edge determined by the modulated pulse and its trailing edge by a fixed pulse. The amplitude of the pulse is independent of supply variations because of the stabilized voltage applied to the grid. This width-modulated pulse is passed through the low-pass filter in the anode circuit; the output from this low-pass filter will contain only the original signal with which the pulses were modulated.[11] This signal is finally amplified to the required level (+10 dBm*) in an audio amplifier.

(5) PERFORMANCE

This equipment was designed to meet the C.C.I.F. requirements for telephone circuits over coaxial cables.[8] Although these requirements are not always directly applicable, they provide a useful standard for comparison with other systems since no recommendations exist for this type of system. Under this heading come noise, crosstalk, distortion and stability.

The input test level to the channels is −14 dBm, and overloading takes place 6 dB above this. The output level is +10 dBm, but this can be varied by adjusting the gain of the output amplifier in 1 dB steps over the range +5 to +15 dBm.

(5.1) Noise

The noise in the output derives from two sources—the radio equipment and the pulse equipment. The radio noise has been mentioned earlier. The C.C.I.F. state that the noise due to all terminal equipment should not exceed 2 500 $\mu\mu$W at a point of zero relative level for a 2 500 km circuit; this corresponds to a signal/noise ratio of 56 dB.

The measured signal/noise ratio with the pulse transmitter and receiver back to back varied from channel to channel in the range 65–75 dB. For values worse than 75 dB the noise is almost entirely due to the use of germanium diodes as limiters to the audio inputs. If a very large number of modulations and demodulations were required, the use of thermionic diodes might be necessary, but since in this type of equipment channels may be extracted and added at will, without demodulating other channels, this is unlikely. The total noise power is limited to 10 000 $\mu\mu$W over a 2 500 km circuit, and this figure is met with the radio equipment used and an average repeater spacing of 30 miles.

(5.2) Crosstalk

The C.C.I.F. require that the intelligible crosstalk between channels be more than 65 dB below the signal.

Measurements showed that the radio equipment used introduced no measurable crosstalk.

With the pulse terminals back to back this crosstalk is, in all cases, more than 70 dB below the signal, and in many cases not measurable.

* Decibels referred to 1 mW.

(5.3) Distortion

The harmonic distortion introduced by the terminals is negligible for all input levels at which limiting does not take place, the total being of the order of 1%. As limiting commences the distortion rises sharply.

In addition to harmonic distortion, outputs are also obtained at frequencies f_r-2f_m, f_r-3f_m, etc.,[11] where f_r is the repetition frequency and f_m is the modulating frequency. The components f_r-3f_m, f_r-4f_m, etc., are negligible. In the case of f_r-2f_m this lies in the band only if f_m is greater than 2 300 c/s. For speech via a normal telephone, components at and above this frequency are relatively negligible, so, although the distortion product is only 36 dB below the modulating signal, the effect is insignificant. This figure for the distortion is well in agreement with theory.

(5.4) Stability

The short-term stability of the equipment in relation to variations of high-voltage and heater supplies due to mains variations is best measured in terms of variations in the overall gain of the various channels.

For a supply variation of ±10% the maximum variation in overall gain is 0·2 dB.

The long-term stability of the pulse terminals is good, and very little adjustment is needed after a few thousand hours' running, apart from occasional valve failures.

(5.5) Summary of Performance

Input test level	−14 dBm (four wire).
Output test level	+10 dBm (four wire).
Overload level	6 dB above test level.
Frequency response ..	±0·5 dB from 300 c/s to 3 400 c/s.
Harmonic distortion ..	Less than 5% or an input 6 dB above test level.
Stability of overall gain ..	Not more than 0·2 dB variation for ±10% variation in h.t. and heater supplies. ±0·5 dB in period of a few weeks.
Cross-talk ratio	Better than 70 dB.
Noise	Better than 65 dB below test level.

(6) CONCLUSIONS

A transmission system of a type suitable for integration with international trunk telephone networks has been described. It has both advantages and disadvantages when compared with the more conventional carrier equipment using frequency-division multiplex. It is less reliable, since very many more valves are used. However, it is very much less bulky than the corresponding frequency-division equipment. It can be used with advantage over types of country where the laying of cable is difficult or its maintenance costly. Because it is less bulky and transmission is by radio, it is relatively more mobile and hence of value for temporary links.

(7) REFERENCES

(1) British Patent No. 150037.
(2) British Patent No. 289102.
(3) United States Patent No. 2051734.
(4) British Patent No. 509820.

(5) JAMES, E. G., DIX, J. C., COPE, J. E., ELLIS, C. F., and ANDERSON, E. W.: "The Development of the Wireless Set No. 10: An Early Application of Pulse-Length Modulation," *Journal I.E.E.*, 1947, **94**, Part IIIA, p. 517.
(6) REID, D. G.: "A 60 cm Multi-Channel System employing Pulse Phase Modulation," *ibid.*, p. 573.
(7) GRIEG and LEVINE: "Pulse-Time Modulated Multiplex Radio Relay System, Terminal Equipment," *Electrical Communication*, 1946, **23**, p. 159.
(8) C.C.I.F. Yellow Books: 1951, **3 bis**, pp. 101, 109 and 121.
(9) CROSBY, M. G.: "Frequency-Modulation Noise Characteristics," *Proceedings of the Institute of Radio Engineers*, 1937, **25**, p. 472.
(10) WENDT, K. R., and FREDENDALL, G. L.: "Automatic Phase and Frequency Synchronization in Television Receivers," *ibid.*, 1943, **31**, p. 7.
(11) FITCH, E.: "The Spectrum of Modulated Pulses," *Journal I.E.E.*, 1947, **94**, Part IIIA, p. 556.
(12) FELDMAN, C. B., and BENNETT, W. R.: "Bandwidth and Transmission Performance," *Bell System Technical Journal*, 1949, **28**, p. 490.
(13) British Patent Application No. 5449/53.

(8) APPENDIX

(8.1) Signal/Noise Advantage in a Pulse-Time-Modulation/Frequency-Modulation System

Feldman and Bennett[12] show that, in a system in which the radio carrier is frequency-modulated by a pulse-time-modulated signal, the mean-square time deviation of a pulse due to noise is given by

$$\overline{\Delta t^2} = \frac{4P_n f_b}{3\pi^2 P_c \beta^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where P_n = Mean fluctuating noise-power per megacycle per second of bandwidth.
f_b = Width of base band (video band), Mc/s.
β = Peak to peak frequency swing of frequency-modulated system, Mc/s.
P_c = Carrier power.

If N = Number of pulse-time-modulated channels,
f_r = Sampling rate, Mc/s,

and maximum depth of modulation is 70%.

$$\text{Then maximum time modulation} = \pm \frac{0\cdot35}{Nf_r} \quad . \quad . \quad . \quad . \quad (2)$$

$$\text{Thus r.m.s. value of signal} = \frac{0\cdot35}{\sqrt{(2)}Nf_r} \quad . \quad . \quad . \quad (3)$$

$$\text{so that} \quad \left(\frac{\text{signal}}{\text{noise}}\right)^2 = \frac{1}{2}\left(\frac{0\cdot35}{Nf_r}\right)^2 \frac{3\pi^2 P_c \beta^2}{4P_n f_b} \quad . \quad . \quad . \quad (4)$$

This value of signal/noise ratio is a voltage ratio. It may just as conveniently be expressed as a power ratio and put in the form

$$\left(\frac{\text{signal}}{\text{noise}}\right)_{\text{power}} = \frac{3}{8}\left(\frac{\pi^2\beta^2 B}{f_b}\right)\left(\frac{0\cdot35}{Nf_r}\right)^2\left(\frac{P_c}{BP_n}\right) \quad . \quad . \quad (5)$$

where B = Radio-frequency bandwidth, Mc/s.

P_c/BP_n represents the ratio of the signal/noise power in the radio receiver; hence the remaining terms constitute the signal/noise advantage.

$$\text{Signal/noise advantage} = \frac{3}{8}\left(\frac{\pi^2\beta^2 B}{f_b}\right)\left(\frac{0\cdot35}{Nf_r}\right)^2 \quad . \quad (6)$$

In this case $B = 12$ Mc/s
$B = 16$ Mc/s
$f_b = 2$ Mc/s
$N = 26$
$f_r = 0\cdot008$ Mc/s

whence the combined pulse-time modulation and frequency-modulation improvements result in a value of 40·8 dB for the ratio of the radio carrier/noise value to the signal/noise ratio in the demodulated channels.

621.372.512

Paper No. 1833 R
May 1955

COUPLED TRANSMISSION LINES AS SYMMETRICAL DIRECTIONAL COUPLERS

By G. D. MONTEATH, B.Sc., Associate Member.

(*The paper was first received* 7*th October*, 1954, *and in revised form* 18*th January*, 1955.)

SUMMARY

If two similar unbalanced transmission lines are coupled together by sharing a common outer conductor for part of their length, so as to form a 3-conductor line, they constitute a symmetrical directional coupler. The output ratio of this varies sinusoidally with frequency, being greatest for lengths equal to an odd number of quarter-wavelengths; the characteristic impedance is independent of frequency, being determined by the cross-section of the 3-conductor line.

Two forms of the directional coupler are described. One of these, intended for laboratory power measurement at frequencies up to 1 000 Mc/s, consists of two parallel strips enclosed in a square-section outer conductor. The other, a very simple device for monitoring transmitter power, consists of two lead-covered coaxial cables grafted together.

Among possible uses suggested are all-pass filters and applications to aerial systems consisting of two parts fed in phase quadrature.

LIST OF PRINCIPAL SYMBOLS

A_1, A_2, B_1, B_2 = Four pairs of input terminals of a symmetrical directional coupler.
C = Capacitance in Maxwell bridge, farads.
c = Velocity of light, m/sec.
K = Output ratio of directional coupler.
$k = \sqrt{(R_u/R_b)}$.
L = Inductance in Maxwell bridge, henrys.
R_0 = Characteristic impedance of directional coupler, ohms.
R_0' = Characteristic impedance of cable from which the simple directional coupler has been made.
R_b, R_u = Characteristic impedance of the 3-conductor line excited respectively as a balanced and an unbalanced two-conductor line (Fig. 3), ohms.
βl = Angular length of 3-conductor line, radians.
θ = Angle subtended by the slot at the axis of either inner conductor of the simple directional coupler, radians.
ω = Angular frequency, rad/sec.

(1) INTRODUCTION

A directional coupler is a device for sampling waves travelling in one direction along a transmission line or waveguide. (Waveguides will not be considered in the paper.) In its more general form it has two pairs of terminals for connection to the transmission line and two auxiliary outputs. When the auxiliary outputs are correctly terminated in resistive loads, a fraction of the energy travelling in one direction along the transmission line is diverted into one load, and the same fraction of the energy travelling in the opposite direction is diverted into the other load. It follows that if the loads are replaced by detectors of the appropriate impedance, a simultaneous indication is obtained of the powers flowing in both directions. The net power transfer may then be obtained as the difference of the powers; their ratio is equal to the square of the reflection coefficient at the load in which the transmission line is terminated.

For many applications it is essential that a directional coupler should not cause reflection in the transmission line into which it is inserted (this will be termed the main transmission line), and that the two auxiliary outputs should be well matched to transmission lines of the same characteristic impedance. The simplest method of satisfying these requirements simultaneously, particularly when close coupling to the main transmission line is required, is to use a symmetrical directional coupler in which there is no distinction between the terminals intended for connection to the main transmission line and those of the auxiliary outputs.

It is convenient to represent a symmetrical directional coupler by the symbol shown in Fig. 1(*a*), in which A_1, A_2, B_1, B_2 each

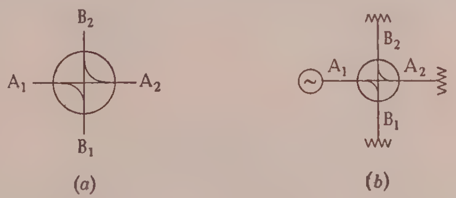


Fig. 1.—(*a*) Symbol representing a symmetrical directional coupler.
(*b*) Action of a directional coupler.

represent a pair of terminals suitable for connection to a transmission line. As suggested by the symbol there are two planes of symmetry, so that each pair of terminals is indistinguishable from any other. In Fig. 1(*b*) three of the pairs of terminals are terminated in pure resistances equal to the characteristic impedance of the transmission line for which the directional coupler has been designed (this will be termed the characteristic impedance of the directional coupler), and the fourth, A_1, is connected to a generator. Part (usually the greater part) of the r.f. power passes along the straight line A_1A_2 into the load at A_2, but a proportion of it passes along the curved line A_1B_1 into the auxiliary load at B_1. If the ratio of the power supplied to the load at B_1 to that supplied to the load at A_2 is denoted by K^2, K may be termed the "output ratio" of the directional coupler. If the directional coupler is perfect and if the loads at A_2 and B_1 are perfectly matched to its characteristic impedance, no power is supplied to the load at B_2. Imperfection of the directional coupler in this respect may be expressed in terms of the directivity (the ratio of the output at B_1 to that at B_2) which is usually expressed in decibels.

A symmetrical directional coupler having perfect directivity has the following properties:[1]

(*a*) Referring to Fig. 1(*b*), the input impedance at A_1 is perfectly matched, provided that the loads at A_2 and B_1 are matched. By symmetry it follows that if the source impedance of the generator at A_1 is matched, the source impedances of the outputs at A_2 and B_1 are matched.

(*b*) The outputs at A_2 and B_1 are in phase quadrature (in view of the symmetry it is not necessary to specify the reference planes of the terminals A_2 and B_1 in making this statement).

The principal object of the work described in the paper was to extend the range of power-measuring equipment for frequencies between 100 and 1 000 Mc/s. For this purpose a range of

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Mr. Monteath is with the British Broadcasting Corporation.

directional couplers was required having various output ratios from about $\frac{1}{3}$ (-10dB) downwards. A minimum directivity of about 30dB was required, and like other components of the equipment, the directional couplers were to be matched to a standing-wave ratio of not less than 0·9. It was clearly desirable to adopt a symmetrical design if possible, since the matching would then be a necessary consequence of the directivity, instead of being a separate requirement. The directional coupler produced is described in Section 4. A simpler form, which has been incorporated in transmitters for monitoring power, is described in Section 5.

(2) TYPES OF DIRECTIONAL COUPLER

(2.1) The Maxwell Bridge

Most transmission-line directional couplers behave at sufficiently low frequencies like the Maxwell bridge illustrated in Fig. 2(a), which has been rearranged in Fig. 2(b) to correspond

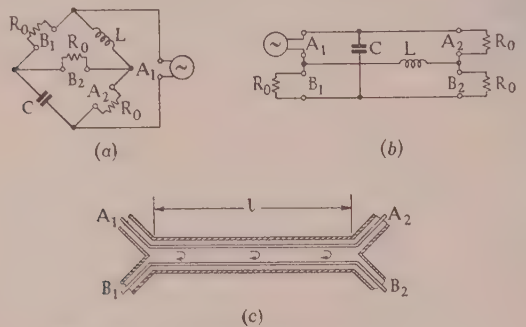


Fig. 2.—(a) Maxwell bridge.
(b) Maxwell bridge circuit rearranged as a directional coupler.
(c) Coupled transmission lines (arrows show direction of coupling).

to the physical configuration of practical directional couplers. The two circuits A_1A_2 and B_1B_2 are coupled by the inductance L and the capacitance C.

Provided that $L/C = R_0^2$ the bridge is balanced at all frequencies, and it therefore behaves as a symmetrical directional coupler of characteristic impedance R_0. The output ratio is given by

$$K = \omega L/R_0 = \omega CR_0 \quad . \quad . \quad . \quad . \quad . \quad (1)$$

The fact that the output ratio is proportional to the frequency may be inconvenient for power measurement, since it accentuates harmonics, whereas usually the fundamental power is required.

The simple circuit of Figs. 2(a) and 2(b) becomes modified by stray inductances and capacitances at high frequencies, particularly if strong coupling (output ratios greater than about -30dB) is required, since the dimensions become comparable with the wavelength. Inductance in series with C may be compensated for by capacitance in parallel with L, but other stray reactances add to the difficulty of achieving high directivity over a wide band of frequencies.

(2.2) The Reflectometer

One of the best-known types of directional coupler is the reflectometer,[2,3] which consists of a small loop coupled to a coaxial feeder so that capacitance and mutual inductance between the loop and the inner conductor perform the functions of C and L in Fig. 2(b). One of the load resistors [at B_1 or B_2 in Fig. 2(b)] is usually built in, so that only one auxiliary output is available. The reflectometer cannot readily be made to give an output ratio greater than 0·05 (-26dB), but if a lower output ratio can be accepted, directivities up to at least 60dB are possible. It is not symmetrical.

(2.3) The Bethe-Hole Directional Coupler

This is a symmetrical directional coupler consisting of two coaxial lines coupled through a common hole in their outer conductors. The diversion of the current path round the hole provides the common inductance L [Fig. 2(b)], while the capacitance between the inner conductors through the hole corresponds to C. Ginzton and Goodwin[5] have obtained output ratios up to 0·1, but the diameter of each coaxial line was then about one-third of the wavelength; the dimensions would be inconveniently large except at microwave frequencies.

(2.4) Coupled Transmission Lines

The directional coupler forming the subject of the paper consists of two unbalanced transmission lines, which are coupled together by sharing a common outer conductor [Fig. 2(c)] so as to form, in effect, a 3-conductor transmission line. It is convenient to ensure symmetry by making the two inner conductors identical. This arrangement may be regarded as a development of the reflectometer; one of the two inner conductors corresponds to the inner conductor of the main transmission line, and the other to the coupling loop. Alternatively it may be considered as a Bethe-hole directional coupler in which the hole has been extended longitudinally. It will be shown that the 3-conductor line acts as a directional coupler even when its length is comparable with the wavelength, and that it has other advantages.

Firestone[6] has investigated directional couplers in which balanced transmission lines consisting of thin wires are coupled together, and has pointed out their relationship to the 3-conductor line, which is the unbalanced equivalent. A more general treatment of balanced coupled transmission lines has recently been given by Oliver.[7]

(3) THEORETICAL CONSIDERATIONS

Figs. 3(a) and 3(b) show a symmetrical 3-conductor transmission line excited as a balanced and an unbalanced 2-conductor line, respectively. Let the characteristic impedance be $2R_b$ for

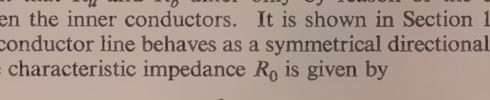


Fig. 3.—Three-conductor line.
(a) As a screened balanced transmission line: characteristic impedance, $2R_b$.
(b) As an unbalanced transmission line: characteristic impedance, $\frac{1}{2}R_u$.

the balanced line and $\frac{1}{2}R_u$ for the unbalanced line. It will be evident that R_u and R_b differ only by reason of the coupling between the inner conductors. It is shown in Section 11.1 that the 3-conductor line behaves as a symmetrical directional coupler whose characteristic impedance R_0 is given by

$$R_0^2 = R_uR_b \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

provided that the velocity of propagation is the same both for the unbalanced and the balanced modes of excitation shown in Fig. 3. This last condition might not be satisfied if the dielectric were not uniform.

It is convenient, in view of eqn. (2), to express R_u and R_b as

$$R_u = kR_0$$
$$R_b = R_0/k \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where k is a parameter which must be greater than unity; its

departure from unity is a measure of the coupling between the inner conductors. The output ratio K is shown in Section 11.1 to be given by

$$K = \tfrac{1}{2}(k - 1/k) \sin \beta l \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where βl is the angular length. An equivalent analysis of balanced coupled transmission lines has been published independently by Oliver.[7] The parameter ζ in Oliver's paper corresponds to k in this paper.

At low frequencies, such that $\beta l \ll 1$, the mutual inductance and capacitance between the inner conductors may be regarded as lumped rather than distributed, so that the coupled transmission lines behave like the Maxwell bridge shown in Fig. 3. The output ratio may be deduced from eqn. (1), if L and C are taken to be the mutual inductance and capacitance between the inner conductors. The result is identical to that deduced from eqn. (4) if $\sin \beta l$ is replaced by βl. At higher frequencies the output ratio varies sinusoidally with frequency; a qualitative explanation of this effect is outlined below.

Referring to Fig. 2(c), the direction of coupling is such that if power flows through the directional coupler from A_1 to A_2, a fraction is diverted to B_1. Each short element of the 3-conductor line may be considered as containing a directional coupler of the Maxwell-bridge type, and the total output at B_1 is the sum of the contributions from different elements. These contributions will take the paths indicated by the arrows in Fig. 2(c), and will arrive at B_1 in different phases. If the length is increased from zero the output ratio will increase linearly until the mis-phasing begins to take effect. When the length reaches one half-wavelength, all phases from 0 to 2π will be represented equally, and the output ratio will be zero.

The maximum output ratio, which is obtained for a length equal to one-quarter wavelength, is equal to $\tfrac{1}{2}(k - 1/k)$. It follows that for a high output ratio k must be large; the inner conductors must be close together so as to form a balanced transmission line of low characteristic impedance, while the overall transverse dimensions of the two inner conductors together must be small compared with the outer conductor, so as to form an unbalanced line of high characteristic impedance when connected as in Fig. 3(b).

(4) A DIRECTIONAL COUPLER FOR GENERAL LABORATORY USE

This Section is concerned with the prototype of a range of directional couplers for power measurement and general laboratory use at frequencies up to 1 000 Mc/s. The principal application envisaged will be discussed in Section 7.1; for the present it is sufficient to list the following requirements:

(a) The directional couplers were to be symmetrical, and to be fitted with four identical r.f. connectors suitable for connection to 71-ohm coaxial cable (Uniradio 21).

(b) The directivity was to be at least 30 dB.

(c) A range of output ratios up to about −10 dB was required.

In order to simplify the design and construction of a range of directional couplers having different output ratios at the same frequency (or giving the same output ratio at different frequencies), it was decided that the directional couplers were to differ only in the length of the 3-conductor line. The cross-section of this line and the configuration of the junction at each end were to be the same in all cases.

Fig. 4 shows the prototype. The 3-conductor line consists of two parallel strips in a square-section outer shield. At each end the two inner conductors are supported by a polystyrene insulator, and are splayed outwards and tapered down to join the inner conductors of matched r.f. connectors.

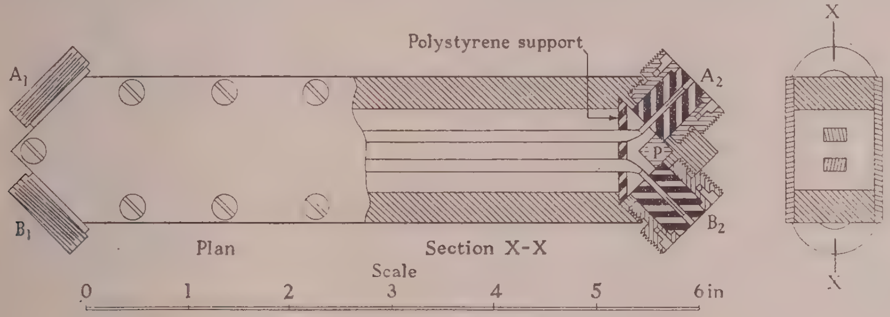


Fig. 4.—Directional coupler for laboratory use.

In the first place it was necessary to choose the cross-section of the 3-conductor line so as to satisfy eqn. (2) and to design the ends so as to ensure that discontinuities and stray coupling would not impair the directivity. The characteristics of the 3-conductor line were determined by means of inductance measurements at 1 Mc/s—a frequency low enough to enable capacitances to be ignored but high enough for skin and proximity effects to be substantially complete. The internal impedance of the conductors, which is negligible at very-high frequencies, could not be neglected completely at 1 Mc/s. A small correction was therefore made for its effect on the measured inductance by observing the measured resistance and using the fact that, when skin effect is substantially complete, the internal resistance and internal reactance are equal. This procedure was applied in the measurement both of self-inductance per unit length and mutual-inductance per unit length.

The 1 Mc/s measurements were made with a medium- and long-wave series impedance bridge* measuring inductance and resistance down to 0·001 μH and 0·01 ohm respectively, and provided with a pair of floating potential terminals. Measurements were made by connecting both inner conductors to the outer at one end, and passing current along one inner conductor from the current terminals of the bridge. Holes drilled through the upper cover-plate enabled a coaxial probe, which was connected by a screened cable to the potential terminals, to be used to sample the potential difference between inner and outer conductors. By measuring this at two points on each inner conductor, the self-inductance and mutual-inductance per unit length were obtained, and R_u and R_b were deduced. Changes were then made in the cross-section until $\sqrt{(R_u R_b)}$ was as near to 71 ohms as could be measured. With the final cross-section the parameter $k = \sqrt{(R_u/R_b)}$ was 1·46. This value indicated a maximum output ratio of 0·39 (−8·2 dB).

After determining the required cross-sectional dimensions of the 3-conductor transmission line in the manner described above, the directional coupler was assembled in the form shown in

* This was developed by C. G. Mayo.

Fig. 4, and measurements of the directivity were made at frequencies in the range 100–900 Mc/s. It was assumed that imperfect directivity could be attributed to discontinuities and stray coupling at the ends, and adjustments were therefore made at the ends to obtain maximum directivity, while preserving symmetry. The directivity was initially determined by observing the apparent reflection coefficient of a 71-ohm resistor fitted into a r.f. connector, but final adjustment and testing could not be carried out in this way because the matching of the load resistors available could not be relied upon to the accuracy required. (Referring to Fig. 1(*b*), if the load at B_2 is replaced by a detector, the observed signal will be the vector sum of components due to reflection at A_2 and B_1 in addition to the component due to imperfect directivity of the directional coupler.) The load resistors were therefore replaced by equal lengths of cable terminated in resistors that were deliberately mismatched, and measurements were carried out over a band of frequencies. Owing to the lengths of cable, the comparatively strong reflections occurring at the resistors varied in phase relative to the apparent reflection introduced by the imperfect directivity. As a result the measured reflection coefficient showed a cyclical variation whose amplitude was a measure of the imperfect directivity.

The end regions in which adjustment is required may be considered as additional short sections of 3-conductor line. In order that these should not impair the directivity they must satisfy the same conditions as the main 3-conductor line. The electrical length must be the same both for balanced and unbalanced modes of propagation, and the characteristic impedances must satisfy eqn. (2). The value of the parameter k applicable to the end sections is unimportant. For example, it could be the same as for the main 3-conductor line, which would in effect be lengthened. Alternatively it would be unity, when the end sections would not contribute to the coupling, but merely act as short lengths of matched coaxial line between the directional coupler and its terminals. Since two independent conditions must be satisfied, two independent adjustments will generally be necessary. For example, these could be equal-value variable capacitors between the inner conductors and the outer, and a variable capacitor between the inners. In fact, the procedure adopted was to design the directional coupler with excessive inductance at each end, and to begin the adjustment by adding the square-section brass blocks marked P in Fig. 4. These were long enough to make contact with both cover-plates, and their cross-section was chosen experimentally for the highest directivity. It happened by chance that the optimum size gave an adequate directivity, and no further change was made.

In the absence of signal generators having accurate attenuators and covering a sufficiently wide band of frequencies, the output ratio was measured with a thermistor milliwattmeter. An oscillator was connected to A_1 through a matched dissipative attenuator, and the powers available at A_2 and B_1 were compared by means of the milliwattmeter, outputs not otherwise connected being terminated in matched load resistors. To reduce the range of powers to be measured, a 10 dB dissipative attenuator, whose attenuation had been measured in a separate experiment, was used to reduce the output from A_2. In Fig. 5 the measured output ratio is plotted against frequency and compared with a sine curve chosen for best fit to the measured points.

Fig. 5 shows that the output ratio satisfies a sinusoidal law within the limits of experimental error. It attains a maximum value at 565 Mc/s, at which frequency one quarter-wavelength is 13·3 cm. This effective length is 0·15 cm longer than the 3-conductor line, measured between the outer surfaces of the polystyrene supports. The maximum output ratio is 0·42 (−7·5 dB), as compared with the value 0·39 deduced from the

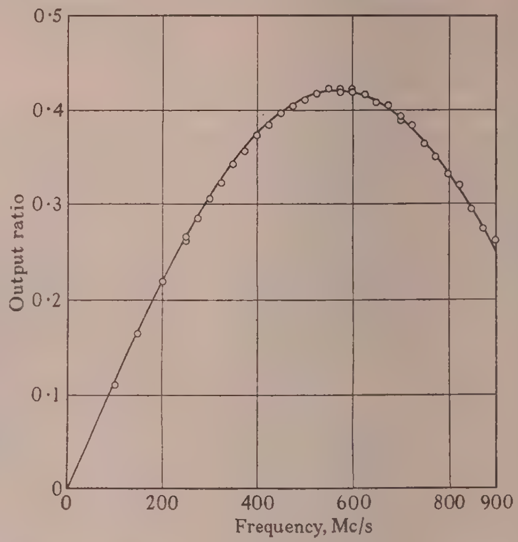


Fig. 5.—Output ratio of directional coupler shown in Fig. 4.
○ Measured values.
——— Sine curve chosen for best fit.

low-frequency inductance measurements. The discrepancy is attributed to error in measuring the small mutual inductance between the inner conductors.

The directivity, measured with the lengths of cable used for the adjustment, exceeded 35 dB at all frequencies in the band 100–900 Mc/s. A lower directivity would probably be observed if different samples of cable were used, owing to the non-uniformity of the characteristic impedance.

(5) A SIMPLE DIRECTIONAL COUPLER FOR MONITORING TRANSMITTER POWER

A different type of directional coupler was required for monitoring transmitter output power in connection with propagation experiments in the v.h.f. and u.h.f. bands. Output ratios of the order of −25 dB were required. A high directivity was not essential, 20 dB being adequate.

Reflectometers could have been designed to give the required performance at the fundamental frequency, but they would have been unduly responsive to harmonics, since the output ratio is proportional to the frequency. This difficulty was overcome by using a 3-conductor transmission line one quarter-wavelength long at the fundamental frequency, which would suppress the even harmonics [$\sin \beta l = 0$ in eqn. (4)], and would couple to the odd harmonics only to the same extent as to the fundamental.

The design of the 3-conductor line was suggested by the fact that if two transmission lines, each having a characteristic impedance R_0' in the absence of the other, are arranged parallel to one another with weak coupling between them, they form a directional coupler whose characteristic impedance is approximately equal to R_0'. Firestone[6] investigated balanced open-wire directional couplers based on this principle, and suggested that their unbalanced equivalents could be devised.

Fig. 6 shows one of the directional couplers. Two lead-covered polythene-filled cables (Uniradio 25) were each milled to form a flat surface, the cut extending into the polythene dielectric, which was exposed in a slot. They were then clamped together with the slots coincidental, and the outer conductors were soldered together. The slots were necessarily tapered at each end where the cables were bent away from the flat surface,

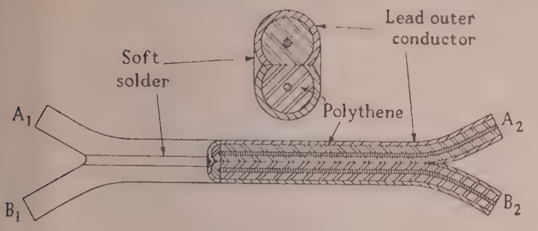


Fig. 6.—Simple directional coupler.

but this was in any case desirable, since the discontinuities caused by rectangular ends would have impaired the directivity. The length of the slot was made slightly greater than one quarter-wavelength, in order to allow for the tapering. One of the cables was arranged to replace an existing cable link in the transmitter; the ends of the other cable were terminated in matched crystal detectors for monitoring the forward and reflected powers.

Section 11.2 contains an approximate theoretical analysis based upon the assumption that the inner conductors are thin compared with their spacing. The result, as applied to 71-ohm polythene-dielectric cables, is illustrated in Fig. 7, which shows the output ratio and the directivity as a function of θ, the angle subtended by the slot at the axis of either inner conductor.

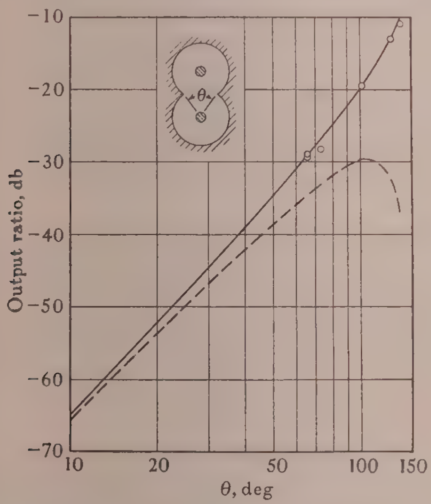


Fig. 7.—Characteristics of directional couplers made from 71-ohm polythene cable.

——— Theoretical maximum output ratio.
○ Measured maximum output ratios.
– – – – Theoretical reciprocal of directivity.

The finite directivity shown in Fig. 7 is a measure of the departure of R_0, which is equal to $\sqrt{(R_u R_b)}$, from R_0', the characteristic impedance of the cable. If θ is increased from zero, R_0 increases from R_0' to a maximum value, the directivity falling from infinity to a minimum. If θ is increased further, R_0 decreases; as it passes through R_0' the directivity again becomes infinite. For values of θ greater than about 100°, at which the minimum directivity occurs, the assumptions upon which the approximate analysis depends are no longer valid, but this qualitative description of the variation of directivity with θ will nevertheless be correct. In any case it would be impracticable to make θ close to 180°, since the inner conductors would be so close together that small mechanical imperfections could have a serious effect.

For small values of θ the output ratio and the reciprocal of the directivity are equal and proportional to θ^2.

A number of directional couplers of the type shown in Fig. 6 have been made for frequencies between 180 and 600 Mc/s. Accurate measurements of the directivity have not been made, but its value exceeds 20 dB in all cases. The maximum output ratios are plotted against θ in Fig. 7; they show good agreement with the theoretical curve.

(6) CASCADE CONNECTION OF SYMMETRICAL DIRECTIONAL COUPLERS

A feature of the directional couplers made from coupled transmission lines is the fact that a high output ratio can be achieved while preserving a high directivity over a wide band of frequencies. Nevertheless practical difficulties may be encountered if output ratios in excess of -10 dB are required, particularly if the frequency and the power to be handled are both high. For example, if the characteristic impedance is 70 ohms and the maximum output ratio is to be unity (0 dB), from eqns. (3) and (4) it is found that R_u and R_b must be 169 and 29 ohms, respectively. These values could be realized by the use of two parallel strips 1 in wide spaced approximately 0·2 in apart in a cylindrical outer conductor approximately 2·5 in in diameter. The narrow gap between the inner conductors would limit the power-handling capacity, but an attempt to overcome this difficulty by increasing the cross-sectional dimensions might impair the directivity at frequencies above about 100 Mc/s owing to end effects. One solution is to obtain a high output ratio by combining together two or more directional couplers of lower output ratio.

Fig. 8(*a*) shows two directional couplers $A_1A_2B_1B_2$ and $A_1'A_2'B_1'B_2'$ of the same characteristic impedance R_0 connected

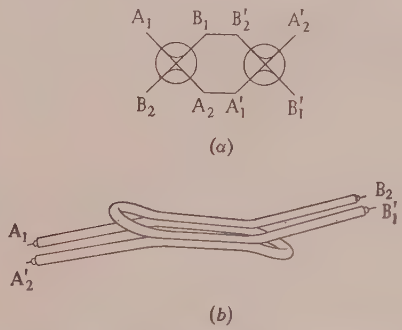


Fig. 8.—Cascade connection of directional couplers.

(*a*) Diagrammatic.
(*b*) A practical arrangement.

together by equal transmission lines A_2A_1' and B_1B_2' to form a single network $A_1A_2'B_1'B_2$. Fig. 8(*b*) shows this method of combination applied to the simple directional couplers described in Section 5. In order to verify that the combination is also a true directional coupler whose characteristic impedance is R_0, suppose that A_2', B_1', and B_2 are each terminated in a resistance R_0 and that a signal is injected at A_1. It is clear that the only possible paths for the signal are $A_1A_2A_1'A_2'$, $A_1A_2A_1'B_1'$, $A_1B_1B_2'B_1'$, and $A_1B_1B_2'A_2'$, and that there can be no output at B_2 and no reflection back towards A_1. Let the input at A_1 be such that the signal reaching A_2' by the path $A_1A_2A_1'A_2'$ is represented in magnitude and phase by unity. Then signals reaching B_1' by the paths $A_1A_2A_1'B_1'$ and $A_1B_1B_2'B_1'$ will respectively be represented by jK' and jK, where K and K' are, respectively, the output ratios of the first and second directional couplers. A signal represented by $-KK'$ will also reach A_2'.

It follows that the total outputs at B_1' and A_2' are respectively $j(K + K')$ and $(1 - KK')$. The combination therefore constitutes a directional coupler of output ratio K_0 given by

$$K_0 = (K + K')/(1 - KK')$$

or $$\arc \tan K_0 = \arctan K + \arctan K' \quad . \quad . \quad . \quad (5)$$

The lengths of the connecting transmission lines B_1B_2' and A_2A_1' are immaterial, provided that they are equal and matched.

Eqn. (5) may be generalized by induction to give the output ratio of a directional coupler formed by the cascade connection of any number n of directional couplers having output ratios $K_1, K_2 \ldots K_r \ldots K_n$. It is given by

$$\arctan K_0 = \sum_{r=1}^{n} \arctan K_r \quad . \quad . \quad . \quad . \quad (6)$$

The band of frequencies in which the output ratio remains within given limits is slightly narrower for a combination of directional couplers of equal length than for a single directional coupler. For example, suppose that the maximum output ratio is unity, and that the bandwidth is determined by the frequencies at which the output ratio falls to 0·9. A single directional coupler would have a bandwidth equal to 57% of the mid-band frequency. Alternatively if three identical directional couplers, each having a maximum output ratio of 0·27 (= tan 15°) were combined, the bandwidth would be reduced to 46% of the mid-band frequency.

A considerable improvement in the constancy of output ratio with frequency may be achieved by combining in cascade directional couplers formed by coupled transmission lines of differing lengths; in this way the bandwidth can be made greater than that of a single directional coupler. The simplest arrangement is to use one or more directional couplers a quarter-wavelength long and one three-quarter-wavelength long at the mid-band frequency; more elaborate combinations are possible.

A method of cascade connection, alternative to that shown in Fig. 8, exists in which B_2 is connected to B_1' and A_2 to A_1'. This method appears to be less useful, since the output ratio depends upon the length of the connecting cables and on the phase delay in the constituent directional couplers. Nevertheless, a pair of coupled transmission lines may be considered as an assembly of elementary Maxwell bridges combined in this way.

(7) APPLICATIONS

(7.1) Power Measurement

Directional couplers are particularly useful for extending the range of r.f. power-measuring devices, such as thermistor milliwattmeters. The most convenient level for power measurement is a few milliwatts, since the thermal time-constant of the thermistor can be made short, and since its physical dimensions can be made small so as to facilitate matching. Lower powers can be measured if steps are taken to compensate for changes in ambient temperature. At frequencies up to 1 000 Mc/s at least, dissipative attenuators can be constructed from rod-and-disc resistors so as to extend the range to powers of the order of 1 watt, but for higher powers the difficulty of matching attenuators increases, since larger resistors are necessary.

A method of using a directional coupler to extend the range of a milliwattmeter to higher powers is indicated in Fig. 9. Although a load capable of absorbing the power to be measured is required, this need not be accurately matched. For many purposes reflection at the load may be ignored, since the error caused by doing so is proportional to the square of the reflection coefficient, but if it must be taken into account, the reflected power may be determined by interchanging the milliwattmeter

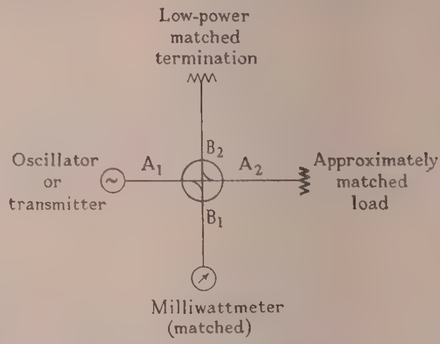


Fig. 9.—Power measurement.

and the low-power matched termination; the net power supplied to the load is then obtained as the difference between the incident and reflected powers. The output ratio of the directional coupler must be determined in a subsidiary experiment, e.g. by using a signal generator having an accurate piston attenuator.

The arrangement of Fig. 9 is particularly convenient when it is required to radiate a known power from an aerial for calibrating field-strength measuring equipment, since the radiated power can be measured at intervals without interrupting the calibration. A similar application to transmitter power monitors has been mentioned in Section 5.

It is sometimes required to generate a known power at too low a level for direct measurement, e.g. to compare with the output of a signal generator. For this purpose the circuit of Fig. 9 may be modified by connecting the milliwattmeter to the directional coupler at A_2 instead of B_1, and taking the low-level output from B_1.

(7.2) All-Pass Networks

An all-pass or constant-resistance network is a lossless four-terminal network whose image impedances are equal resistances independent of the frequency. When this is connected between a source and a load that are matched to the image impedance, the insertion loss is zero at all frequencies, although the phase shift is in general a non-linear function of the frequency. Since all-pass networks are not networks of minimum phase shift, there must always be at least two paths between the input and output.

Various all-pass networks may be made up from directional couplers and lengths of lossless transmission line. Two examples are shown in Figs. 10(*a*) and 10(*b*). The all-pass character of these networks may be verified by tracing out the path of the signal as indicated by the arrows, and noting that if the output is terminated in a matched load no energy is reflected back to the input. The insertion loss is therefore zero if dissipation can be neglected. More complicated networks can be constructed by inserting subsidiary all-pass networks in the closed loop of transmission line.

Fig. 10(*c*) shows the simplest unbalanced all-pass network, which corresponds to the network of Fig. 10(*b*) with the length of the loop of transmission line reduced to zero. The phase shift ϕ may be shown to be given by

$$\tan(\phi/2) = (1/k)\tan\theta \quad . \quad . \quad . \quad . \quad (7)$$

where θ is the angular length of the three-conductor line and k is the parameter defined by eqns. (3). The group delay varies with frequency in a cyclical manner about the mean value; it is a maximum and minimum when θ is an odd multiple and an even multiple of $\pi/2$, respectively. The response to a positive "unit-impulse" signal may be shown to consist of a positive

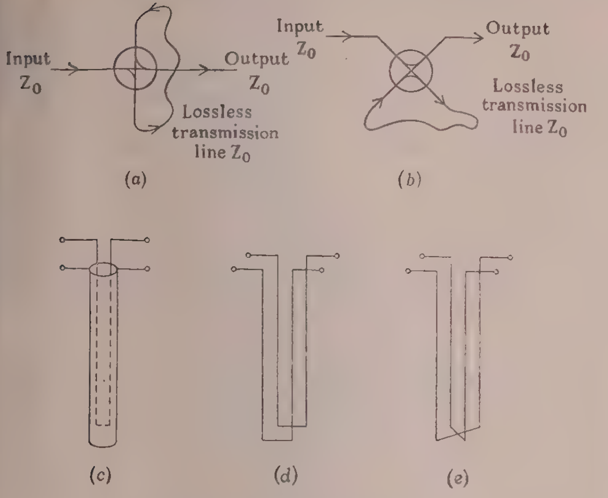


Fig. 10.—All-pass networks.

(*a*), (*b*) Networks consisting of a directional coupler and a length of transmission line.
(*c*) Simplest unbalanced network corresponding to (*b*).
(*d*) Balanced equivalent of (*c*).
(*e*) Network (*d*) modified by transposition of cross-connections.

impulse coincident in time with the input but reduced in amplitude, followed by a train of impulses of alternating sign, the first being positive.

The balanced equivalent of the network of Fig. 10(*c*), which is shown in Fig. 10(*d*), requires no comment. It may be modified by transposing the cross-connections, as shown in Fig. 10(*e*), to form a new all-pass network which has no simple unbalanced equivalent. To obtain the phase shift of this network, $1/k$ in eqn. (7) is replaced by k.

The all-pass networks have been discussed mainly as a curiosity, but applications may exist. For example, they could be used for ensuring a quadrature phase relationship between the feeds to two aerials over a wide band of frequencies. Phillips[8] has employed conventional all-pass networks for this purpose at frequencies up to a few megacycles. An alternative method of maintaining a quadrature phase relationship over a wide band of frequencies is discussed in the following Section.

(7.3) Quadrature Hybrid Networks

Referring to eqn. (4) it will be seen that K attains a maximum value at the frequency at which $\theta = \pi/2$, i.e. when the length of the directional coupler is equal to one quarter-wavelength. At neighbouring frequencies K will be approximately constant. By arranging that the maximum value is equal to unity (or slightly greater than unity) K can be made approximately equal to unity over a wide band. (For example if its maximum value is $1 \cdot 05$, K will lie between $0 \cdot 95$ and $1 \cdot 05$ over a band whose width is 56% of the mid-band frequency.) When K is equal to unity the directional coupler belongs to a class of networks which may be termed *quadrature hybrid networks*. These have applications in aerial systems consisting of two parts fed with equal powers in phase quadrature. Other networks of this type are known, but they are more complicated and do not cover such a wide band.

In one application, illustrated in Fig. 11(*a*), the directional coupler, acting as a quadrature hybrid network, could be used to combine the outputs of two transmitters without the use of frequency-selective networks. For example, it could take the place of the diplexer and a quarter-wavelength of transmission line in an aerial system of the type used at the Sutton Coldfield television station, which has been described by Bevan and Page.[9]

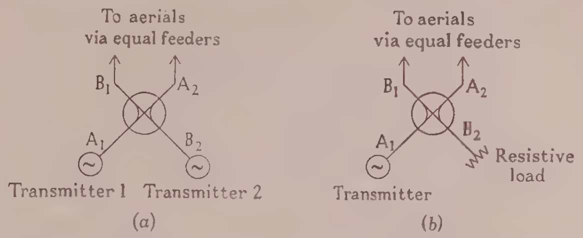


Fig. 11.—Applications to phase-quadrature aerial systems.

(*a*) Diplexer.
(*b*) Power-equalizing network.

Fig. 11(*b*) illustrates a similar application, in which the directional coupler, again acting as a quadrature hybrid network, forms, with a resistive load, a *power-equalizing network*. In this arrangement the impedance of the aerials may be allowed to depart more widely from the characteristic impedance of the feeders than would be permissible if the aerials were fed in parallel through feeders differing by one quarter-wavelength. Provided that the two aerials have the same impedance, departure of this impedance from the characteristic impedance of the feeders does not affect the radiation pattern or the input impedance at A_1, since energy reflected owing to mismatch is absorbed in the resistive load at B_2. The use of a power-equalizing network in this way has been described by Masters.[10] The principal advantage of using the directional coupler (with a resistive load) as a power-equalizing network is the wide band of frequencies covered. A possible application would be the radiation of two television programmes from a single aerial of the "superturnstile" type, which has itself a wide bandwidth.

For many transmitting applications, power-handling difficulties would necessitate the use of directional couplers connected in cascade in order to achieve the required output ratio of unity.

(8) CONCLUSIONS

Directional couplers consisting of coupled transmission lines offer advantages over other types for general laboratory use. A directivity exceeding 35 dB has been achieved at frequencies up to 900 Mc/s with a maximum output ratio of $-7 \cdot 5$ dB. A simple form of directional coupler made from lead-covered cable has given directivities exceeding 20 dB without any form of experimental adjustment. The sinusoidal variation of output ratio with frequency enables a fairly constant output ratio to be obtained over a band of frequencies by arranging the length to be one quarter-wavelength at the mid-band frequency. This procedure also suppresses even harmonics of the mid-band frequency; the output ratio for odd harmonics is then the same as that for the fundamental.

Up to the present the directional couplers described have been used only for power measurement, but the possibility of other uses exist. These include a range of all-pass networks consisting of directional couplers and lengths of transmission line, and applications to aerial systems consisting of two parts fed in phase quadrature.

(9) ACKNOWLEDGMENTS

The author is indebted to the Chief Engineer of the British Broadcasting Corporation for permission to publish the paper. He also wishes to express his appreciation of the help given by Mr. G. H. Millard who carried out the experimental work described in the paper.

(10) REFERENCES

(1) MONTGOMERY, DICKE, and PURCELL: "Principles of Microwave Circuits" (McGraw-Hill, 1948).

(2) EARLEY, H. C.: "A Wide-Band Directional Coupler for Wave Guide," *Proceedings of the Institute of Radio Engineers*, 1946, **34**, p. 883.

(3) ALLAN, H. R., and CURLING, C. D.: "The Reflectometer," *Proceedings I.E.E.*, Paper No. 767 R, January, 1949 (**96**, Part III, p. 25).

(4) SOSIN, B. M.: "Application of Directional Couplers to Amplitude Ratio and Phase Measurement at Very High Frequencies," *Marconi Review*, 1951, **14**, p. 39.

(5) GINZTON, E. L., and GOODWIN, P. S.: "A Note on Coaxial Bethe-Hole Directional Couplers," *Proceedings of the Institute of Radio Engineers*, 1950, **38**, p. 305.

(6) FIRESTONE, W. L.: "Directional Coupling with Transmission Lines," *Tele-Tech*, November, 1953, **12**, No. 11, p. 95.

(7) OLIVER, B. M.: "Directional Electromagnetic Couplers," *Proceedings of the Institute of Radio Engineers*, 1954, **42**, p. 1686.

(8) PHILLIPS, G. J.: "A Wide-Band Aerial System for Circularly Polarized Waves, Suitable for Ionospheric Research," *Proceedings I.E.E.*, Paper No. 1121 R, May, 1951 (**98**, Part III, p. 237).

(9) BEVAN, P. A. T., and PAGE, H.: "The Sutton Coldfield Television Broadcasting Station," *ibid.*, Paper No. 116 R, April, 1951 (**98**, Part III, p. 416).

(10) MASTERS, R. W.: "A Power-Equalizing Network for Antennas," *Proceedings of the Institute of Radio Engineers*, 1949, **37**, p. 735.

(11) APPENDICES

(11.1) Derivation of the Results of Section 3

Fig. 12 shows a 3-conductor transmission line characterized by push-pull and push-push characteristic impedance $2R_b$ and

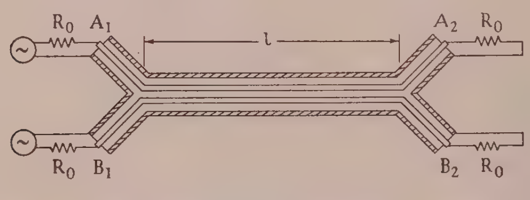


Fig. 12.—Three-conductor line characterized by R_u and R_b (defined in Fig. 3).

R_0 equals the characteristic impedance of the coaxial lines. Source impedance of generators is zero.

$\frac{1}{2}R_u$ as defined in Fig. 3. The inner conductors are connected to coaxial transmission lines terminated in their characteristic impedance R_0, which is assumed to be equal to $(R_bR_u)^{1/2}$. The generators shown connected in series with the loads at A_1 and B_1 are assumed to have zero self-impedance. If the generator at A_1 only is assumed to be operative, this will cause waves to pass along the coaxial line to the 3-conductor line. As a result waves may in general pass from the 3-conductor line towards A_2, B_1 and B_2, and there may also be reflection back to A_1. It is required to find the relative amplitudes and phases of these waves.

The problem is simplified by considering a second generator connected at B_1, as shown in the Figure, and assuming the two generators to operate firstly in phase and secondly in antiphase. By applying the superposition theorem the effect of a single generator at A_1 is obtained. In what follows the amplitudes of waves will be characterized by the voltage, rather than the current.

First it will be assumed that each generator produces a positive impulse, so that a positive unit impulse travels from A_1 and B_1 towards the left-hand junction. Two coaxial lines in parallel have a characteristic impedance $\frac{1}{2}R_0$, and the 3-conductor line behaves like an unbalanced transmission line of characteristic impedance $\frac{1}{2}R_u = \frac{1}{2}kR_0$. The voltage reflection coefficient at the junction is therefore $(k-1)/(k+1)$; positive impulses of this magnitude will be returned to A_1 and B_1. An impulse of magnitude $2k/(k+1)$ will travel from left to right along the 3-conductor line. Since the reflection coefficient at the right-hand junction will be $(1-k)/(k+1)$, an impulse of magnitude $4k/(k+1)^2$ will pass into the loads at A_2 and B_2, and an impulse of magnitude $2k(1-k)/(k+1)^2$ will travel from right to left along the 3-conductor line. The last-mentioned impulse will now be reflected backwards and forwards in the 3-conductor line, and it will be changed in sign at each reflection. A geometrically-decreasing train of positive impulses will arrive at A_2 and B_2, and a train of negative impulses will arrive at A_1 and B_1.

The result may be summarized as follows. If waves characterized by the unit impulse $\delta(t)$ are incident on the left-hand junction from A_1 and B_1, the resulting output at A_2 and B_2 is characterized by the function $H_1(t)$, where

$$H_1(t) = \frac{4k}{(k+1)^2}\sum_0^\infty\left(\frac{k-1}{k+1}\right)^{2n}\delta[t-(2n+1)l/c] \quad . \quad (8)$$

The waves reflected back to A_1 and B_1 are characterized by the function $H_2(t)$, where

$$H_2(t) = \frac{k-1}{k+1}\delta(t) - \frac{4k}{(k+1)^2}\sum_0^\infty\left(\frac{k-1}{k+1}\right)^{2n+1}\delta[t-(2n+2)l/c] \quad . \quad . \quad . \quad . \quad (9)$$

A similar process occurs if the generator at B_1 is reversed, so that the two generators are excited in anti-phase; the reflection coefficient at either junction is found to be of the same magnitude as before but reversed in sign. The outputs at A_1, A_2, B_1, and B_2 are found to be $-H_2(t)$, $H_1(t)$, $H_2(t)$, and $-H_1(t)$, respectively.

If the outputs due to the push-push and push-pull excitations are added and the sum is halved, the output due to a unit impulse incident on the left-hand junction from A_1 is obtained. It is zero at A_1 and B_2, $H_1(t)$ at A_2, and $H_2(t)$ at B_1.

By expressing the outputs at A_2 and B_1 in terms of their frequency spectra, the ratio of the output at B_1 to that at A_2 at an angular frequency ω is found to be

$$\tfrac{1}{2}j(k-1/k)\sin(\omega l/c)$$

where c is the velocity of light.

The fact that the two outputs are in phase quadrature and that there is no output at B_2 establishes that the device is a directional coupler of characteristic impedance R_0. The output ratio as stated in eqn. (4) is confirmed.

(11.2) Analysis of the Directional Coupler described in Section 5

It is required to find the push-push and push-pull characteristic impedances, $\frac{1}{2}R_u$, $2R_b$, of the 3-conductor line. The geometric mean of these quantities is the characteristic impedance of the directional coupler, whose departure from the characteristic impedance of the coaxial cable from which the directional coupler is made determines the finite directivity. The ratio R_u/R_b determines the output ratio.

The method is to change the cross-section of the 3-conductor line into a simpler form by a conformal transformation. Provided that the transformation is 1 : 1 (this excludes transformations which divide one conductor into two or more, and those which superimpose two or more conductors), R_u and R_b will be unchanged by the transformation.

An approximation will be made by assuming that the radius of the inner conductors is small compared with their spacing from each other and from the outer conductor. Expressed in physical terms the approximation is equivalent to neglecting the non-uniformity of the distribution of current round the periphery of each inner conductor. Mathematically, the procedure is as follows. Suppose that in the z-plane the axis of an inner conductor is represented by the complex number z_1, and that its periphery is the locus $z = z_1 + a\varepsilon^{j\theta}$, a being the radius. If now the transformation $w = \mathrm{f}(z)$ is made, the new axis will be at $w = \mathrm{f}(z_1)$ and the periphery will be the locus $w = \mathrm{f}(z_1 + a\varepsilon^{j\theta})$. If a is small the latter expression may be expanded by Taylor's theorem and powers of a above the first may be neglected:

$$w \simeq \mathrm{f}(z_1) + a\varepsilon^{j\theta}\mathrm{f}'(z_1) \quad . \quad . \quad . \quad . \quad (10)$$

It follows that to a first approximation the periphery in the w-plane remains a circle centred on the point $w = \mathrm{f}(z_1)$ corresponding to the centre in the z-plane. The radius is, however, changed from a to $a|\mathrm{f}'(z_1)|$. It would not be difficult to obtain a second approximation by taking into account the next term of the Taylor series—the effect is mainly to displace the centre of the periphery slightly—but unwieldy expressions result.

The transformation is illustrated in Fig. 13, in which (a) is the cross-section of the 3-conductor line. For convenience the radius of the outer conductor is taken to be unity, and the width of the slot is expressed in terms of θ, the angle subtended at the axis of either inner conductor. The actual cross-section will be regarded as the z-plane.

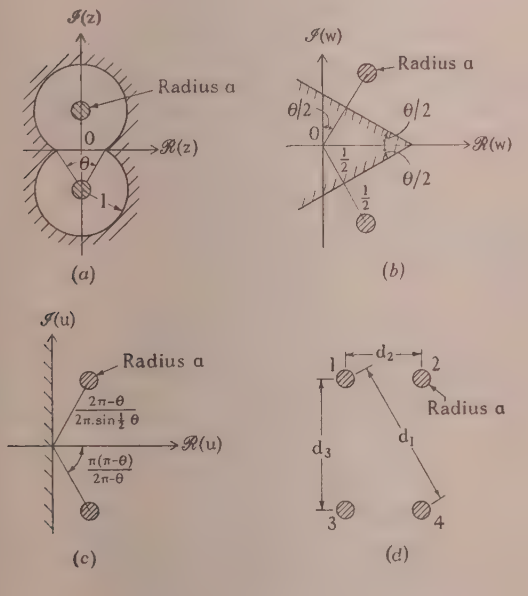


Fig. 13.—Conformal transformation of the cross-section.

(a) z-plane (original cross-section).
(b) w-plane. $w = (z + \sin\tfrac{1}{2}\theta)^{-1}$.
(c) u-plane. $u = \dfrac{2\pi-\theta}{2\pi\sin\frac{1}{2}\theta}(2w\sin\tfrac{1}{2}\theta - 1)^{\pi/(2\pi-\theta)}$
(d) Conducting half-plane replaced by images of conductors in it. *Change of scale.*

$$d_1 = \frac{2\pi-\theta}{\pi\sin\frac{1}{2}\theta}$$

$$d_2 = \frac{2\pi-\theta}{\pi\sin\frac{1}{2}\theta}\sin\frac{\pi\theta}{4\pi-2\theta}$$

$$d_3 = \frac{2\pi-\theta}{\pi\sin\frac{1}{2}\theta}\cos\frac{\pi\theta}{4\pi-2\theta}$$

The first stage in the transformation is to carry out a geometrical inversion with respect to the left-hand edge of the slot, the radius of inversion being unity. This is done by putting

$$w = (z + \sin\tfrac{1}{2}\theta)^{-1} \quad . \quad . \quad . \quad . \quad . \quad (11)$$

The resulting cross-section in the w-plane is shown in Fig. 13(b). It will be seen that the outer conductor now takes the simpler form of a conducting wedge, which partially screens the two inner conductors from one another. If the width of the slot tends to zero in Fig. 13(a), the wedge in Fig. 13(b) becomes more acute and its apex moves to the right, until eventually the two inner conductors are completely separated by an infinite conducting plane. The radius of each inner conductor is unchanged, since $|dw/dz|$ is unity at the axis.

The cross-section in the w-plane is further simplified by moving the origin to the apex of the wedge and raising w to a fractional power so as to reduce the angle of the sector not occupied by the wedge from $2\pi-\theta$ to π. At the same time the dimensions are simplified by making a change of scale. This is done by putting

$$u = \frac{2\pi-\theta}{2\pi\sin\frac{1}{2}\theta}(2w\sin\tfrac{1}{2}\theta - 1)^{\pi/(2\pi-\theta)} \quad . \quad . \quad (12)$$

The derivative of u is given by

$$\frac{du}{dw} = (2w\sin\tfrac{1}{2}\theta - 1)^{-(\pi-\theta)/(2\pi-\theta)} \quad . \quad . \quad . \quad (13)$$

At the axis of either inner conductor $|du/dw|$ is equal to unity; the radius of the inner conductors therefore remains unchanged. The resulting cross-section in the u-plane is shown in Fig. 13(c).

Instead of considering the cross-section of Fig. 13(c) directly, the conducting plane into which the outer conductor has finally been transformed is replaced by the images of the inner conductors in it, as shown in Fig. 13(d). We now have to consider a balanced directional coupler formed by two open-wire transmission lines coupled together.

Referring to Fig. 13(d), R_u is equal to the characteristic impedance of the balanced transmission line formed by connecting conductors 1 and 3 and conductors 2 and 4. R_b is equal to the characteristic impedance of the balanced transmission line formed by connecting conductors 1 and 4 and conductors 2 and 3. Assuming that the radius of the inner conductors is small compared with the spacing between them the result is found to be

$$R_u = 60\log_\varepsilon\frac{d_1 d_2}{a d_3}$$

$$= 60\log_\varepsilon\left(\frac{2\pi-\theta}{a\pi\sin\frac{1}{2}\theta}\sin\frac{\pi\theta}{4\pi-2\theta}\right) - 60\log_\varepsilon\cos\frac{\pi\theta}{4\pi-2\theta} \quad (14)$$

$$R_b = 60\log_\varepsilon\frac{d_2 d_3}{a d_1}$$

$$= 60\log_\varepsilon\left(\frac{2\pi-\theta}{a\pi\sin\frac{1}{2}\theta}\sin\frac{\pi\theta}{4\pi-2\theta}\right) + 60\log_\varepsilon\cos\frac{\pi\theta}{4\pi-2\theta} \quad (15)$$

Eqns. (14) and (15) may be interpreted as follows. The first term in the expressions for R_u and R_b is the characteristic impedance of the unbalanced transmission line formed by one inner conductor and the outer conductor, the other inner conductor being removed. If θ tends to zero this term tends to R_0', the characteristic impedance of the coaxial cable from which the directional coupler has been made. This is given by

$$R_0' = 60\log_\varepsilon(1/a) \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

For finite but small values of θ the presence of the slot increases the characteristic impedance by an amount proportional to θ^2. The second term takes account of the coupling between the inner conductors through the slot; it is positive in the expression for R_u and negative in the expression for R_b, and is proportional to θ^2 for small values of θ.

The output ratio K and the characteristic impedance R_0 of a directional coupler constructed from the 3-conductor line of Fig. 13(a) are obtained by substituting for R_u and R_b in eqns. (4) and (2), respectively. In practice it is assumed in using the directional coupler that its characteristic impedance is equal to R_0', which is given by eqn. (16). It can be shown that if the difference between R_0 and R_0' is small, as is the case in practice, the directivity is equal to

$$\frac{R_0'\sqrt{(1+K^2)}}{|R_0 - R_0'|}$$

In most applications the square root in this expression will be approximately equal to unity.

DIGESTS OF INSTITUTION MONOGRAPHS

THE RADIATION OF A HERTZIAN DIPOLE OVER A COATED CONDUCTOR

621.396.674.3 Monograph No. 113 R

D. B. BRICK, Ph.D., S.M., A.B.

(DIGEST *of a paper published in December*, 1954, *as an* INSTITUTION MONOGRAPH *and republished in March*, 1955, *in Part C of the* PROCEEDINGS.)

This paper treats the problem of an electric or magnetic dipole in or above a uniform layer of dielectric covering an infinite, perfectly-conducting plane. The results show that a dipole near or in the dielectric layer is a strong source for surface waves of the Attwood[4] type. These are waves which are totally internally reflected within the dielectric region.

Two types of waves exist in the region above the dielectric—the spherical- or radiated-type wave and one or more surface guided-type waves. The latter are waves of cylindrical type which attenuate exponentially with height and which propagate with phase velocity less than that of light in the half-space above the dielectric surface.

For the special cases of vertical electric dipoles in or near fairly thin dielectric coatings the far-zone fields have been combined at a typical distance for a polystyrene dielectric layer to yield field patterns. Several examples of these are given illustrating the distortion of the dipole fields due to the coating on the conductor. To further illustrate this distortion, expressions were derived for the total power output of the dipole, the ratio of the powers in the guided and radiated waves, and the radiation resistances of the dipoles as functions of dipole position. Several curves of these quantities, with comparison curves for metal surfaces without dielectric coatings, are given. In addition, the expressions for the attenuation coefficients of the guided waves due to finite but large conductivities of the conducting plane are derived by means of a perturbation method.

No attempt was made to compute the attenuation due to dielectric loss or to describe the field and flux distributions of the surface waves. These have been amply treated by Attwood.[4]

LIST OF PRINCIPAL SYMBOLS

$R, \theta\, \Phi$ = Spherical co-ordinates, m, rad, rad.
r, Φ, z = Cylindrical co-ordinates, m, rad, m.
d = Dielectric thickness, m.
z' = Dipole height, m.
n^2 = Dielectric constant of dielectric layer relative to that of the half-space above the layer.
ϵ, μ = Dielectric constant and permeability of half-space above dielectric layer, farads/m, henrys/m.
$\omega = 2\pi \times$ frequency, rad/sec.
$\beta = \omega\sqrt{(\epsilon\mu)}$ = Characteristic propagation coefficient of the half-space, m^{-1}.
λ_j = Radial propagation coefficient of the jth surface mode, m^{-1}.
p = Dipole moment, coulomb-m.
m = Magnetic dipole moment, amp-m^2.
$M(\theta)$ = θ-component of the electric-field relative to its maximum value for fixed value of R.
P = Time average power flow, watts.
α_j = Attenuation coefficient of the jth surface mode, nepers/m.
σ = Conductivity, mhos/m.
R^e = Radiation resistance, ohms.
Superscript r denotes the radiated wave.
Superscript G denotes the guided wave.

Dr. Brick is in the Division of Applied Science, Harvard University.

FORMULATION

The configuration of interest is the infinite region shown in Fig. 1, with an infinitesimal electric Hertzian dipole, a magnetic dipole, or an Abraham dipole (half a Hertzian dipole erected

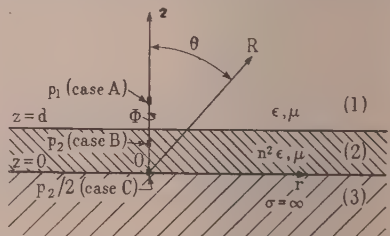


Fig. 1.—Cross-section of the coated perfect conductor showing the co-ordinate system used.

on an image plane which with its image forms a complete dipole) oriented along and parallel to the z-axis of the cylindrical co-ordinate system r, Φ, z. Region 1 is a half-space, region 2 is the dielectric region of thickness d with dielectric constant n^2 or refractive index n relative to region 1, and region 3 is a perfect conductor.

The problem is formulated in terms of the z-components of the electric or magnetic Hertz vectors. These satisfy inhomogeneous Helmholtz equations, where the inhomogeneous terms are of the form of constants times Dirac delta functions. The

olutions containing unintegrated terms are readily determined n terms of cylindrical co-ordinates through the application of he Fourier–Bessel or Hankel transform[7] pair for cylindrical ymmetry to the Helmholtz equations the boundary conditions eing derived from Maxwell's equation and the radiation conlition at infinity.[5,6] Solutions are obtained for the fields both n and above the dielectric layer.

The results have the form of an infinite complex integral, arts of which could be evaluated immediately. The evaluation of these parts yields spherical waves with source points at the dipole positions and at their images in the dielectric or metal urfaces. The forms of the integrands of the remaining unintegrated terms show that a continuous spectrum of waves, multiply reflected from the dielectric and metal surfaces, exists.

The integrands of these remaining terms are of the type which llow asymptotic integrations of the classical[10,11,12] or of the Van der Waerden[18] saddle-point methods to be utilized. The lectric Hertz potentials in regions 1 and 2 owing to vertical lectric dipoles in or fairly near thin dielectric coatings are valuated in this manner.

These integrands have first-order poles at solutions, λ_j, of the ollowing transcendental equation:

$$n^2\sqrt{(\lambda_j^2 - \beta^2)} = \sqrt{(n^2\beta^2 - \lambda_j^2)} \tan [d\sqrt{(n^2\beta^2 - \lambda_j^2)}] \quad . \quad (1)$$

The roots of this equation must satisfy

$$\beta \leqslant \lambda_j \leqslant n\beta$$

As βd increases, each λ_j increases from β towards $n\beta$, and the quation has an increasing number of solutions. An additional olution occurs whenever βd attains the value

$$\beta d = \frac{k\pi}{\sqrt{(n^2 - 1)}}, \text{ where } k = 0, 1, 2, \ldots$$

The residues evaluated at the poles, determined by solutions f eqn. (1), are of the form of cylindrical waves, which propagate n the r-direction with propagation constants λ_j or with phase elocity ω/λ_j, which is less than ω/β, and which attenuate as xp $[- z\sqrt{(\lambda_j^2 - \beta^2)}]$ in region 1 (see Fig. 1). In region 2 the aves have z-dependence cos $[z\sqrt{(n^2\beta^2 - \lambda_j^2)}]$ These waves hug" the surface and propagate with velocity less than that haracteristic of region 1. Hence they are called guided, surface, r slow waves. They also result as eigenfunctions of the conguration[4] (see Fig. 1) (excluding $r = 0$). The remaining terms, ncluding the parts which could be immediately integrated, were esignated as the radiating or compensating waves.

Upon carrying out the saddle-point integrations, it was shown at the compensating waves were of the spherical type.

The Far-Zone E- and B-Fields

The representations for the far-zone E- and B-fields are btained by applying the well-known[3] relations to the Hertz otentials. The same two types of waves exist.

The guided modes are elliptically polarized in region 1 with mi-major axis in the z-direction and semi-minor axis along r. he eccentricity for each mode, which is independent of position, β/λ_j.

Figs. 2–4 contain some examples of the computed field patterns otted on a semi-logarithmic scale. The field patterns are rves of the relative magnitudes (with respect to their maxima) the θ-components of the electric fields in region 1, M(θ), as nctions of θ, in the range of θ between 0 and $\pi/2$, for $\beta R \gg 1$. he dielectric is assumed to be thin, so that $\theta \simeq \pi/2$ along the electric surface. The value of $\beta R = 239$ was arbitrarily chosen, d the dielectric constant ($n^2 = 2\cdot54$) of polystyrene was used

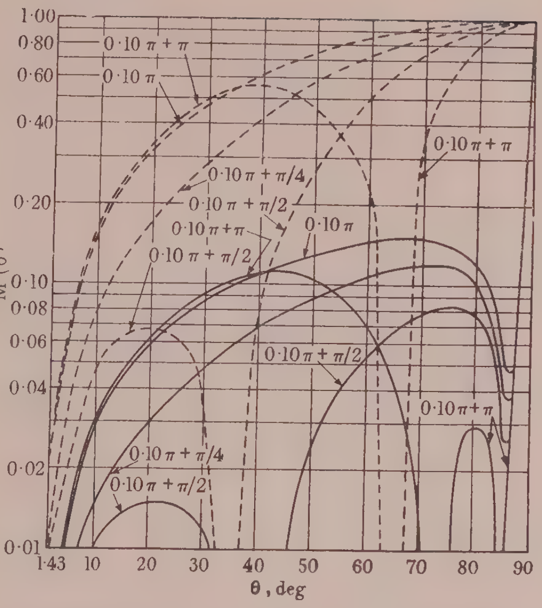


Fig. 2.—Field patterns for case A ($\beta d = 0\cdot10\pi$) and comparative field patterns for $\beta d = 0$.

$\beta R_0 \simeq 239$; $n^2 = 2\cdot54$.
The curves, which are marked accordingly, are plotted for $\beta z' = 0\cdot10\pi$, $\pi/4 + 0\cdot10\pi$; $\pi/2 + 0\cdot10\pi$; and $\pi + 0\cdot10\pi$.
——— $\beta d = 0\cdot10\pi$; — — — — $\beta d = 0$.

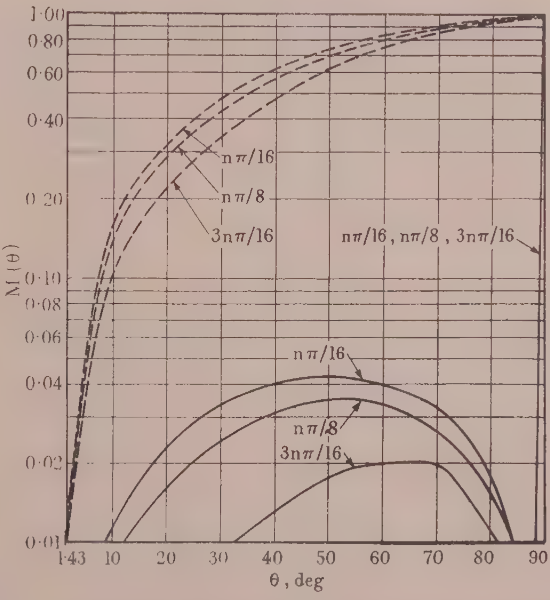


Fig. 3.—Field patterns for case B ($\beta d = n\pi/4$) and comparative patterns for $\beta d = 0$.

$\beta R_0 \simeq 239$; $n^2 = 2\cdot54$.
The curves, which are marked accordingly, are plotted for $\beta z' = n\pi/16$; $n\pi/8$; and $3n\pi/16$.
——— $\beta d = n\pi/4$; — — — — $\beta d = 0$.

for the coating. βd is restricted to the region of single-mode propagation in all the computations.

The main points of interest in these field patterns are the large maxima or spikes at $\theta \simeq \pi/2$ and the compressions of the rest of the fields with the accompanying minima between these two.

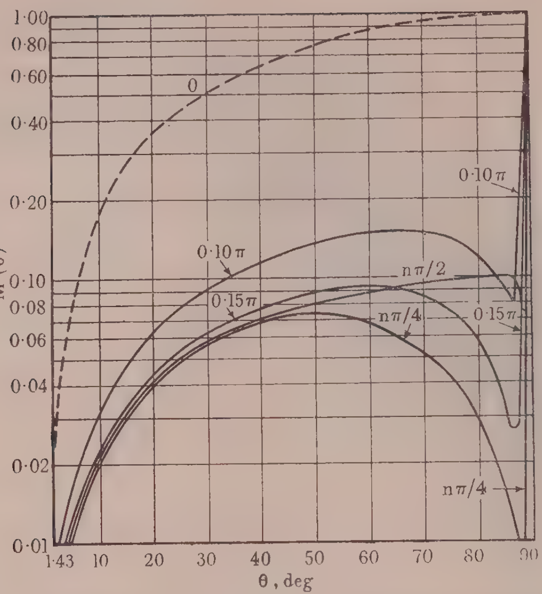


Fig. 4.—Field patterns for case C ($\beta z' = 0$) and comparative patterns for $\beta d = 0$.

$\beta R_0 \simeq 239$; $n^2 = 2 \cdot 54$.
— — — — $\beta d = 0$.
———— $\beta d = 0 \cdot 10\pi$; $0 \cdot 15\pi$; $n\pi/4$; and $n\pi/2$.

Formulation of the Power Relations

The purpose of this Section is to show that the total time average power flowing across a surface surrounding a vertical dipole is composed of two or more independent parts—that due to the radiated wave alone and those due to each of the guided modes.

The total fields for all values of βr greater than zero may be represented in the following way:

$$\boldsymbol{E} = \boldsymbol{E}^r + \sum_{J=1}^{k} \boldsymbol{E}_J^G \quad . \quad . \quad . \quad . \quad . \quad . \quad (2a)$$

$$\boldsymbol{B} = \boldsymbol{B}^r + \sum_{J=1}^{k} \boldsymbol{B}_J^G \quad . \quad . \quad . \quad . \quad . \quad . \quad (2b)$$

where superscript r denotes the radiated wave, superscript G denotes the guided wave, and k is the number of surface modes which can propagate.

The power flowing across a surface S surrounding the source is given by

$$P = \mathscr{R}\left\{\frac{1}{2\mu}\left[\int_S \hat{n} \cdot (\boldsymbol{E}^r \times \boldsymbol{B}^{r*})dS + \sum_{J=1}^{k}\int_S \hat{n} \cdot (\boldsymbol{E}^r \times \boldsymbol{B}_J^{G*} + \boldsymbol{E}_J^G \times \boldsymbol{B}^r)dS + \sum_{J=1}^{k}\sum_{m=1}^{k}\int_S \hat{n} \cdot (\boldsymbol{E}_J^G \times \boldsymbol{B}_m^G)dS\right]\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where superscript * indicates the complex conjugate of the quantity.

It can be proved through the utilization of the reciprocity relationship[21] and specially chosen surfaces, S, that each term of the second integral is zero and that the third integral is zero, for $\lambda_J \neq \lambda_m$, in eqn. (3). In the course of this proof a method is given for deriving the magnitudes of the surface waves for a particular source and dielectric layer without resorting to the evaluation of the residues and consideration of the path of integration.

The total power radiated is then given by the following formula:

$$P = \mathscr{R}\left\{\frac{1}{2\mu}\left[\int_S \hat{n} \cdot (\boldsymbol{E}^r \times \boldsymbol{B}^{r*})dS + \sum_{j=1}^{k}\int_S \hat{n} \cdot (\boldsymbol{E}_j^G \times \boldsymbol{B}_j^{G*})dS\right]\right\} = P^r + P^G \quad . \quad (4)$$

Computations were made of these quantities as functions of the electrical-dipole height, $\beta z'$, for $n^2 = 2 \cdot 54$, and for two thicknesses of dielectric, $\beta d = 0 \cdot 15$ and $n\pi/4$, both in the region of single-mode surface-wave propagation. Fig. 5 contains

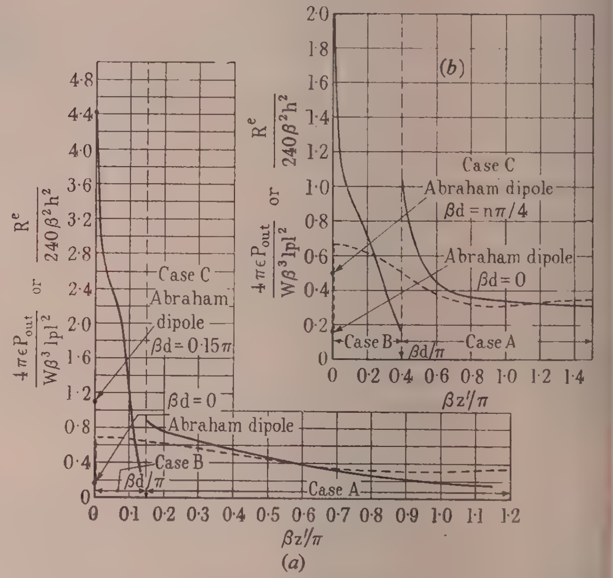


Fig. 5.—Total power radiated and radiation resistance as a function of $\beta z'$.

(a) $n^2 = 2 \cdot 54$.
———— $\beta d = 0 \cdot 15\pi$; – – – – $\beta d = 0$.
(b) $n^2 = 2 \cdot 54$.
———— $\beta d = n\pi/4$; – – – – $\beta d = 0$.

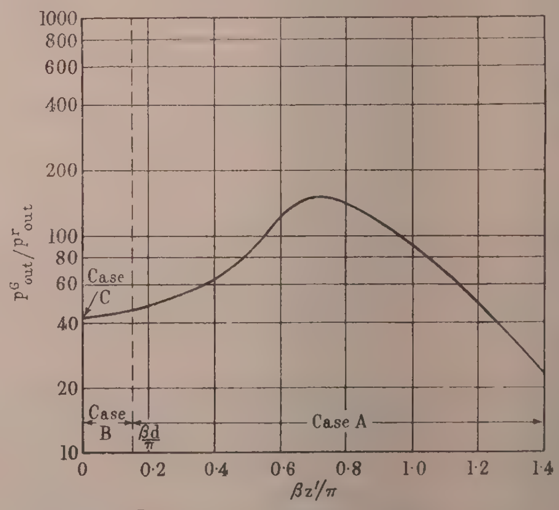


Fig. 6.—P^G_{out}/P^r_{out} as a function of $\beta z'$ for $\beta d = 0 \cdot 15\pi$.

$n^2 = 2 \cdot 54$.
This includes cases A, B and C as $\beta z'$ varies.

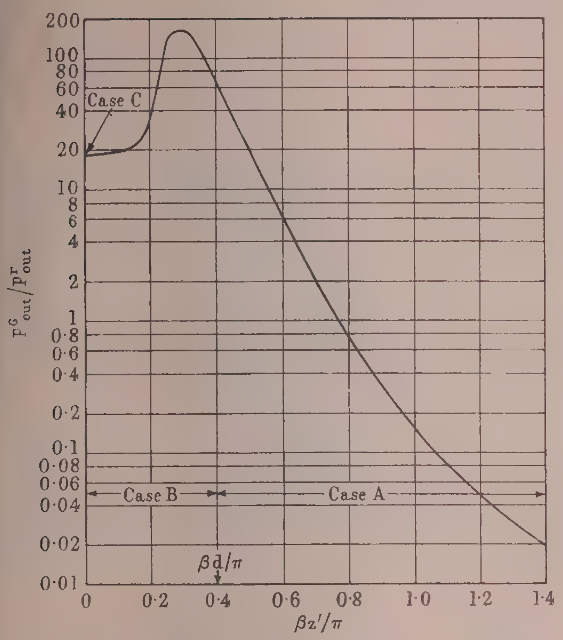


Fig. 7.—P^G_{out}/P^r_{out} as a function of $\beta z'$ for $\beta_d = n\pi/4$.
$n^2 = 2\cdot 54$
This includes cases A, B and C as $\beta z'$ varies.

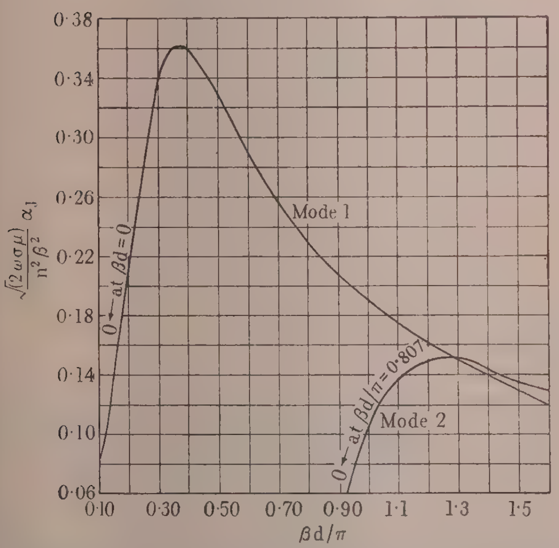


Fig. 8.—Attenuation of a surface mode due to a large but finite conductivity σ of the ground plane as a function of dielectric thickness βd.
$n^2 = 2\cdot 54;\ 0\cdot 10\pi \leqslant \beta d \leqslant 1\cdot 6\pi$.

urves of $4\pi\epsilon P/(|p^2|\omega\beta^3)$, where p is the dipole moment, for the wo values of βd and for $\beta d = 0$ for comparison. The disconinuities at $\beta z' = \beta d$ are due to differences in the magnitudes of he sources in regions 1 and 2.

Figs. 6 and 7 contain curves of P^G/P^r versus dipole height $\beta z'$ or the same βd's and n as in Fig. 5. These curves show the overvhelming proportion of the power that is fed to the surface vave for a dipole near or in the dielectric surface.

Utilizing the perturbation formula the attenuation[25] coefficient, α_J, for $\beta r \gg 1$, of the Jth guided mode due to a large but finite conductivity, σ, of the ground plane is derived. Fig. 8 contains a plot of the quantity $\sqrt{(2w\sigma\mu)}\alpha_J/n^2\beta^2$ as a function βd for $n^2 = 2/54$ for modes 1 and 2.

The radiation resistance[24] of the dipole is given by

$$\frac{R^e}{240\beta^2h^2} = \frac{4\pi\epsilon P}{\omega\beta^3|p|^2} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where h is the half-length of a Hertz dipole or the length of an Abraham dipole. This quantity is plotted in Fig. 5 together with P.

References

(1) Marcuvitz, N.: "Guided Wave Concept in Electromagnetic Theory," *Research Report R*–269–52 (P.I.B.–208, ONR, May 14, 1952. Polytechnic Institute of Brooklyn, Microwave Research Institute), p. 1.
(2) Sommerfeld, A.: "Partial Differential Equations in Physics" (Academic Press, New York, 1949), p. 246.
(3) Stratton, J. A.: "Electromagnetic Theory" (McGraw-Hill, New York, 1941), p. 573.
(4) Attwood, S.: "Surface Wave Propagation over a Coated Plane Conductor," *Journal of Applied Physics*, 1951, **22**, p. 504.
(5) Frank-Mises: "Differentialgleichungen der Physik" (Frederich Vieweg und Sohn, Brunswick, Germany, 1935), Vol. II, Chap, XIX, Sec. 5.
(6) King, R. W. P.: "Electromagnetic Engineering" (McGraw-Hill, New York, 1945), Vol. II, p. 232.
(7) Stratton, J. A.: *loc. cit.*, p. 371.
(8) Stratton, J. A.: *ibid.*, pp. 486–487.
(9) Stratton, J. A.: *ibid.*, pp. 575–576.
(10) Tai, C. T.: "The Effect of a Grounded Slab on Radiation from a Line Source," *Journal of Applied Physics*, 1951, **22**, p. 405.
(11) Stratton, J. A.: *loc. cit.*, p. 584.
(12) Karp and Sollfrey: "Diffraction by a Dielectric Wedge," *Research Report* No. EM–23 (Mathematical Research Group, New York University, October, 1950), pp. 19 and 31.
(13) Van der Waerden, B. L.: "On the Method of Saddle Points," *Applied Scientific Research*, **B–2**, No. 7, p. 43.
(14) Jahnke and Emde: "Tables of Functions" (Dover, New York), p. 134.
(15) King, R. W. P.: *loc. cit.*, pp. 268 and 276.
(16) Goubau, G.: "Surface Waves and Their Application to Transmission Lines," *Journal of Applied Physics*, 1950, **21**, p. 1119.
(17) Goubau, G.: "Single Conductor Surface-Wave Transmission Lines," *Proceedings of the Institute of Radio Engineers*, 1951, **39**, pp. 619–624.
(18) Goubau, G.: "On the Excitation of Surface Waves," *ibid.*, 1952, **40**, p. 865.
(19) Goubau, G.: "Uber die Zennecksche Bodenwelle," *Zeitschrift für Angewandt Physik*, 1951, **3**, p. 103.
(20) King, R. W. P.: *loc. cit.*, p. 220.
(21) Stratton, J. A.: *loc. cit.*, p. 479.
(22) King, R. W. P.: *loc. cit.*, p. 221.
(23) Stratton, J. A.: *loc. cit.*, p. 435.
(24) King, R. W. P.: *loc. cit.*, p. 296.
(25) Lamont, H. R. L.: "Wave Guides" (John Wiley, New York, 1949), p. 45.

THE IONOSPHERIC PROPAGATION OF RADIO WAVES OF FREQUENCY 16kc/s OVER SHORT DISTANCES

621.396.11.029.4 Monograph No. 114 R

T. W. STRAKER, M.Sc., Ph.D.

(Digest *of a paper published in January*, 1955, *as an* Institution Monograph *and republished in March*, 1955, *in Part C of the* Proceedings.)

The paper describes a continuation of the investigations of Best, Budden, Ratcliffe and Wilkes[1,2] on the ionospheric propagation of radio waves of frequency 16kc/s emitted from the Post Office sender GBR at Rugby, and observed at Cambridge, 90km away. This work formed part of a comprehensive programme of radio research at the Cavendish Laboratory on the propagation of long and very long waves in the ionosphere, the results of which have been described in a series of companion papers.[3-10] Although primarily undertaken to test the tentative conclusions of the pre-war work[1,2] and for comparison with the results of companion work at higher frequencies and greater distances, the present long series of detailed observations has revealed some new and interesting effects.

The experiments have been made on continuous-wave transmissions using a method previously developed[1] to isolate the downcoming wave and to make separate measurements on its amplitude and phase relative to the ground wave. In general, the downcoming wave is elliptically polarized and may be resolved into two plane-polarized components—the normal component with its electric vector in the plane of propagation, and the abnormal component with its electric vector perpendicular to this plane. In the present experiments observations were made on the abnormal component alone, and it was assumed that its behaviour represented that of the downcoming wave as a whole. There is strong evidence[1,2,5] to support this assumption.

Observations on signals from GBR were made on 362 days between 11th March, 1948, and 30th September, 1949. At all times a strong downcoming wave was received. On most days the phase and amplitude changed smoothly and regularly, while on others they were more or less violently disturbed. In order to investigate the regular behaviour of the downcoming wave, days which were obviously abnormal were excluded, and the remaining curves were grouped in sets corresponding to 14 days and mean curves were drawn through them. The curves of Figs. 1 and 2 are typical of the average curves obtained in this way.

It is convenient to describe the results in terms of a model in which the ionospheric wave is reflected, possibly with a change of phase, from a sharply defined reflecting surface with a change of amplitude measured by a reflection coefficient. It is not intended to imply that the ionospheric wave is in fact returned to earth in this way. The ratios of the amplitudes of the normal and abnormal components of the wave, reflected from the fictitious surface, to that of the incident wave (which is entirely normal component) are referred to as the reflection coefficient $_{||}R_{||}$ and the conversion coefficient $_{||}R_{\perp}$, respectively. There is some evidence[1,2,5] to indicate that these coefficients are roughly equal for the conditions of the present experiments.

On a few occasions simultaneous observations were made on signals from GBR (16kc/s, 90km) and GBZ (19·4kc/s, 222km). By using an adaptation of the frequency-change method of Appleton and Barnett[11] it was then possible to deduce the apparent height of reflection from the measured difference between the phases of the downcoming waves on the two frequencies. The results for 5th October, 1949, are shown in Fig. 3. On this day there were two phase anomalies of the type associated with solar flares, lasting from 0942h to 1230h.

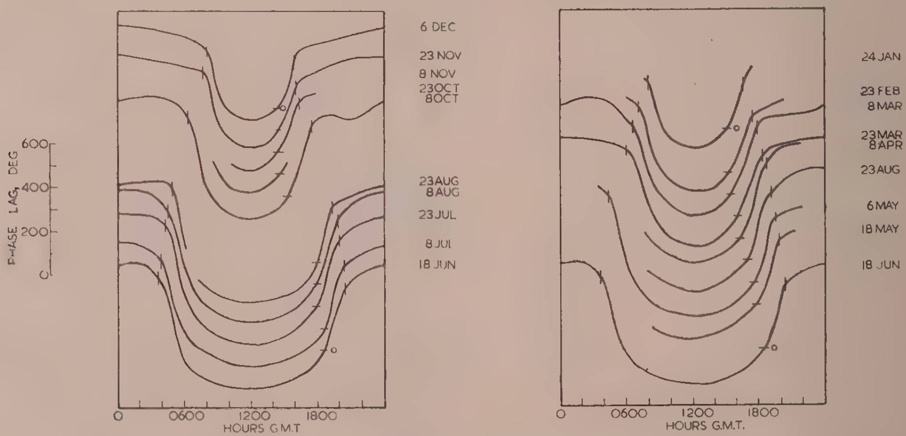


Fig. 1.—Diurnal variation of the phase of the abnormal component in 1948 and 1949.
| Time of sunrise and sunset.
— Zero phase relative to ground wave.

The 14-day average curves of Fig. 1 summarize the regular behaviour of the phase of the downcoming wave. If it is assumed that changes of phase result entirely from changes of height of a fictitious reflecting surface, these curves may be regarded as showing the way in which the apparent height varies throughout the day. All the results may then be represented by an equation of the form

$$h_{\chi} = h_{\chi=0} + A(t) \log \sec \chi \quad . \quad . \quad . \quad . \quad (1)$$

where h_{χ} and $h_{\chi=0}$ represent the apparent heights when the zenith distance of the sun is χ and 0 respectively, and $A(t)$ is a factor which varies slowly with the time of year as shown in Fig. 4(*a*), with a marked semi-annual periodicity. The results shown in Fig. 4(*b*) indicate that $h_{\chi=0}$ is approximately constant, while the measurements of apparent height, previously mentioned, suggest that its value is 69km.

It is interesting to note that if reflection took place from a

Dr. Straker is with the Defence Research Board, Canada.

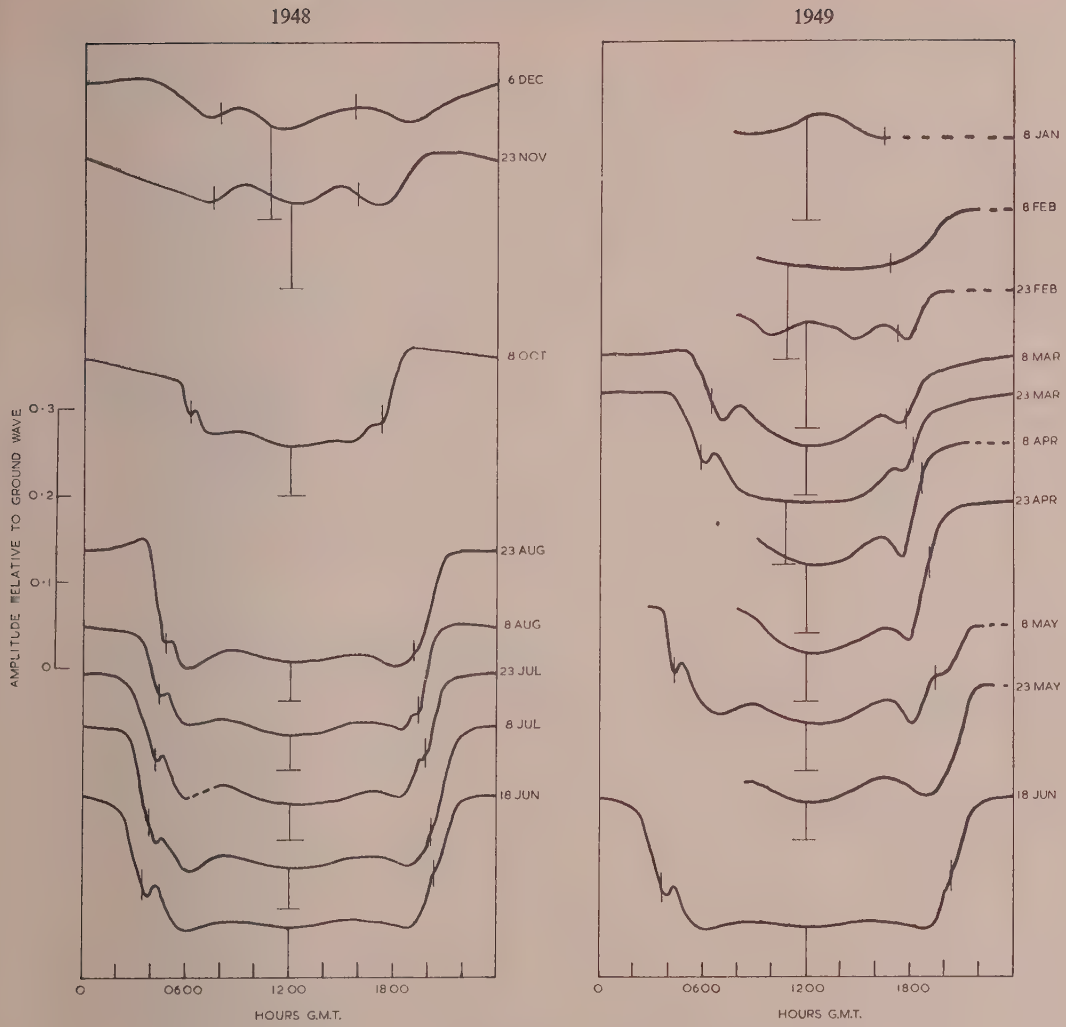


Fig. 2.—Diurnal variation of the amplitude of the abnormal component in 1948 and 1949.
| Times of sunrise and sunset.

level of constant electron density in the lower parts of a Chapman region,[13] $A(t)$ would represent the scale height of the atmosphere.[2] It is not intended to imply that the height of the fictitious reflector does in fact correspond to a height of constant electron density. Only a full-wave theory of the reflection of very low frequency waves can settle this point, and such a theory does not seem to be available at present.

Fig. 5 shows that the seasonal variations of apparent height at local noon and at night are markedly different. At night there is no large semi-annual variation of the type found by day, reflection taking place from about 87 km at all times of the year.

The 14-day average curves of Fig. 2 summarize the regular behaviour of the amplitude of the abnormal component. In summer the amplitude is closely controlled by the sun but is not related in any simple way to its zenith distance, whereas in winter it is substantially constant throughout the day and night. Fig. 6 shows that the seasonal variations of the conversion coefficients by day and by night are markedly different. The rapid increase of the daytime amplitude in October followed by a gradual decrease in the spring is similar to the "November effect" on the strength of low-frequency waves transmitted across the Atlantic.[15,16,17]

It is clear in many ways that the phase and amplitude vary in an independent manner. This evidence suggests the tentative hypothesis that reflection and absorption occur at different levels.

The curves of Figs. 1 and 2 show that the daily cycles of phase change and amplitude change usually start quite suddenly near sunrise, the times of these phenomena being closely controlled by the sun but in different ways. The sunrise effect on the amplitude was clearly evident only from March to October and occurred about 56 min before ground sunrise, whereas at all times of the year the effect on the phase occurred about 9 min after sunrise.

Two distinct types of irregular behaviour have been observed. The first type, which is closely associated with solar flares, has been described in detail elsewhere.[2,4] The second type appears to be associated, but not in a very precise way, with magnetic disturbances.

The most violent and prolonged irregularities appear to be associated with great magnetic storms. The results of detailed observations made during and after the great magnetic storm of 7th August, 1948,[18] are shown in Fig. 7. These reveal some peculiar effects. Apart from some violent fluctuations of the phase and amplitude at night, there were no marked irregularities

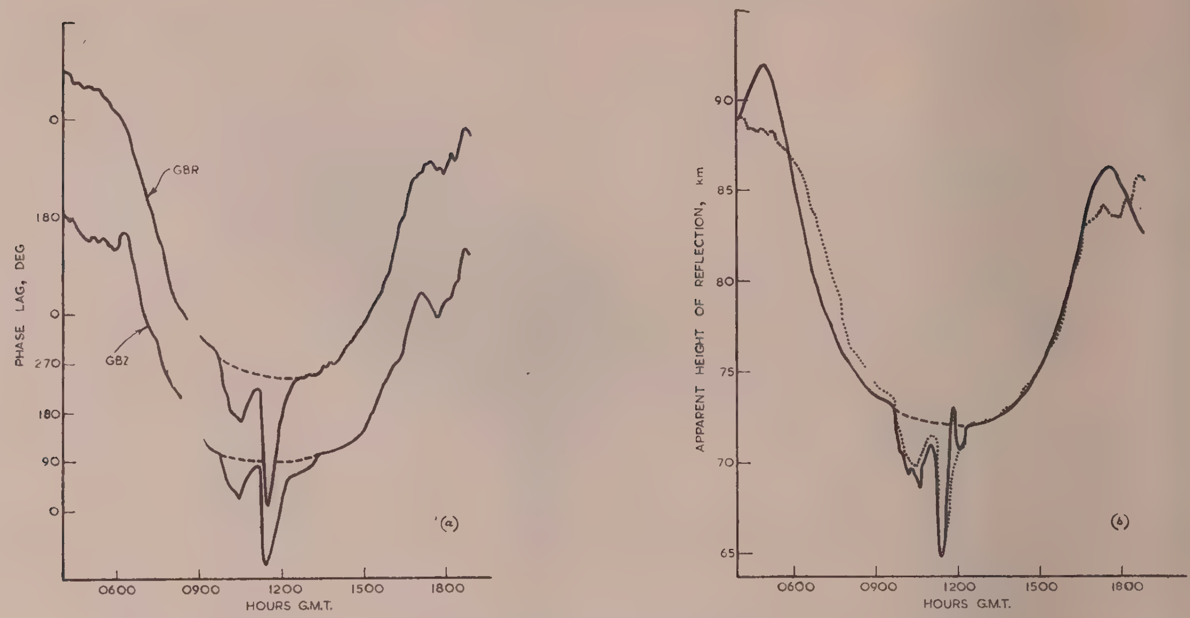


Fig. 3.—Determination of the height of reflection by the frequency-change method, 5th October, 1949.

(*a*) Observed phase of the abnormal component of the downcoming wave from GBR (16kc/s, 90km) and GBZ (19·4kc/s, 222km).
(*b*) Apparent height of reflection deduced from the phase curves (*a*). The dotted curve shows, for comparison, the variation of reflection height deduced from the phase curve for GBR, on the assumption that changes of phase are entirely due to changes in reflection height.

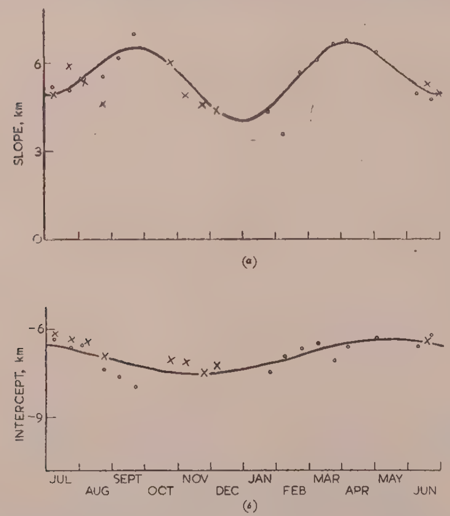


Fig. 4.—Seasonal variation of the parameters $A(t)$ (slope) and $h_{\chi=0}$ (intercept) in the equation $h_\chi = h_{\chi=0} + A(t) \log \sec \chi$.

× Observations made in 1948.
○ Observations made in 1949.

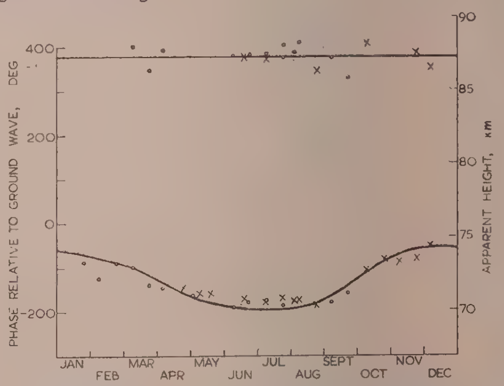


Fig. 5.—Seasonal variation of the apparent height of reflection at night (upper curve) and at local noon (lower curve).

× Observations made in 1948.
○ Observations made in 1949.

during the period of greatest storm activity. The most striking effect occurred three days after the onset of the storm when magnetic activity had subsided to a low level. The phase then remained substantially constant by day and night for three days, resuming its regular daily variation another four days later. Somewhat similar effects have been observed on very-low-frequency transmissions over short[2,6] and long[16] distances.

Anomalies of a less violent nature often appear to be associated with small magnetic storms, but it is abundantly clear that the two phenomena are not related in any simple way. Fig. 8 shows an example of a phase anomaly when there was a clear-cut correspondence between the two phenomena.

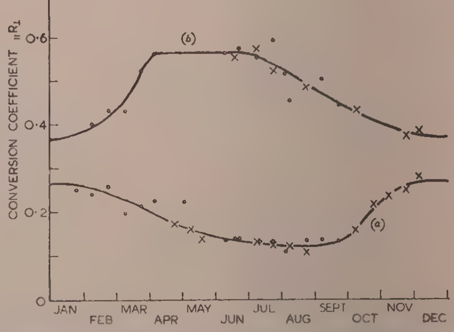


Fig. 6.—Seasonal variation of the conversion coefficient $_{\parallel}R_{\perp}$ by day [curve (*a*)] and by night [curve (*b*)].

× Observations made in 1948.
○ Observations made in 1949.

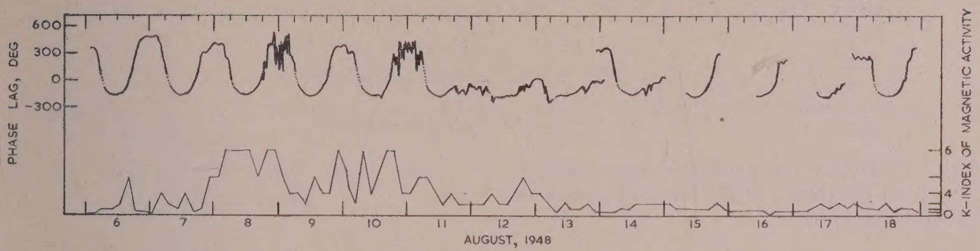


Fig. 7.—Diurnal variation of the phase of the abnormal component and variation of the *K*-index of magnetic activity during and after the great magnetic storm of 7th August, 1948.

REFERENCES

(1) BEST, J. E., RATCLIFFE, J. A., and WILKES, M. V.: "Experimental Investigation of Very Long Waves reflected from the Ionosphere," *Proceedings of the Royal Society*, A, 1936, **156**, p. 614.

(2) BUDDEN, K. G., RATCLIFFE, J. A., and WILKES, M. V.: "Further Investigations of Very Long Waves reflected from the Ionosphere," *ibid.*, 1939, **171**, p. 188.

(3) WEEKES, K.: "The Ground Interference Pattern of Very Low Frequency Waves," *Proceedings I.E.E.*, Paper No. 926 R, March, 1950 (**97** III, p. 100).

(4) BRACEWELL, R. N., and STRAKER, T. W.: "The Study of Solar Flares by Means of Very Long Radio Waves," *Monthly Notices of the Royal Astronomical Society*, 1949, **109**, p. 28.

(5) BRACEWELL, R. N.: "The Ionospheric Propagation of Radio Waves of Frequency 16kc/s over Distances of about 200km," *Proceedings I.E.E.*, Monograph No. 34 R, April, 1952 (**99** IV, p. 217).

(6) BAIN, W. C., BRACEWELL, R. N., STRAKER, T. W., and WESTCOTT, C. H.: "The Ionospheric Propagation of Radio Waves of Frequency 16kc/s over Distances of about 540km," *Proceedings I.E.E.*, Monograph No. 37 R, May, 1952 (**99** IV, p. 250).

(7) BRACEWELL, R. N., HARWOOD, J., and STRAKER, T. W.: "The Ionospheric Propagation of Radio Waves of Frequency 30–65kc/s over Short Distances," *Proceedings I.E.E.*, Monograph No. 84 R, December, 1954 (**101** IV, p. 154.

(8) WEEKES, K., and STUART, R. D.: "The Ionospheric Propagation of Radio Waves of Frequency near 100kc/s over Short Distances," *Proceedings I.E.E.*, Monograph No. 15 R, 1952 (**99** IV, p. 29).

(9) WEEKES, K., and STUART, R. D.: "The Ionospheric Propagation of Radio Waves of Frequency near 100kc/s over Distances up to 1 000km," *Proceedings I.E.E.*, Monograph No. 16 R, November, 1952 (**99** IV, p. 38).

(10) BRACEWELL, R. N., BUDDEN, K. G., RATCLIFFE, J. A., STRAKER, T. W., and WEEKES, K.: "The Ionospheric Propagation of Low and Very Low Frequency Radio Waves over Distances less than 1 000km," *Proceedings I.E.E.*, Paper No. 1123, May, 1951 (**98** III, p. 221).

(11) APPLETON, E. V., and BARNETT, M. A. F.: "On Some Direct Evidence for Downcoming Atmospheric Reflection

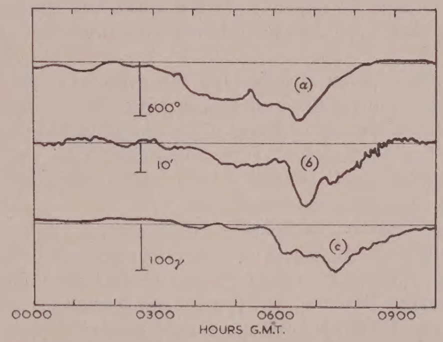


Fig. 8.—Deviation of the phase, compared with the variation of the magnetic field at Abinger, England, and Cheltenham, U.S.A., 10th October, 1948.

(*a*) Deviation of the phase on 16kc/s.
(*b*) Variation of the magnetic declination at Abinger.
(*c*) Variation of the vertical magnetic field at Cheltenham.

of Electric Rays," *Proceedings of the Royal Society*, A, 1925, **109**, p. 621.

(12) BREIT, G., and TUVE, M. A.: "A Test of the Existence of the Conducting Layer," *Physical Review*, 1926, **28**, p. 559.

(13) CHAPMAN, S.: "The Absorption and Dissociative or Ionizing Effect of Monochromatic Radiation in an Atmosphere on a Rotating Earth," *Proceedings of the Physical Society*, 1931, **43**, pp. 26 and 483.

(14) HAVENS, R. J., KOLL, R. T., and LAGOW, H. E.: "The Pressure, Density and Temperature of the Earth's Atmosphere to 160 Kilometers," *Journal of Geophysical Research*, 1952, **57**, p. 59.

(15) ROUND, H. J., ECKERSLEY, T. L., TREMELLEN, K., and LUNNON, F. C.: "Report on Measurements made on Signal Strength at Great Distances during 1922 and 1923 by an Expedition sent to Australia," *Journal I.E.E.*, 1925, **63**, p. 933.

(16) ESPENCHIED, L., ANDERSON, C. N., and BAILEY, A.: "Trans-Atlantic Radio Telephone Transmission," *Proceedings of the Institute of Radio Engineers*, 1926, **14**, p. 7.

(17) HOLLINGWORTH, J.: "The Propagation of Radio Waves," *Journal I.E.E.*, 1926, **64**, p. 579.

(18) "Principal Magnetic Storms," *Terrestrial Magnetism and Atmospheric Electricity*, 1948, **53**, p. 481.

A TABLE OF A FUNCTION USED IN RADIO-PROPAGATION THEORY

621.396.11 : 538.566.2 Monograph No. 115 R

F. HORNER, M.Sc., Associate Member.

[DIGEST *of a paper (official communication from the D.S.I.R. Radio Research Station) published in January,* 1955, *as an* INSTITUTION MONOGRAPH *and republished in March,* 1955, *in Part C of the* PROCEEDINGS.]

For the calculation of the field from an aerial, expressions are used which are based on the theory of Sommerfeld[1] and which involve a factor $F(\omega)$ given by[2,3]

$$F(\omega) = 1 + 2j\sqrt{(\omega)}\varepsilon^{-\omega}\int_{-j\sqrt{\omega}}^{\infty}\varepsilon^{-x^2}dx$$

The parameter ω depends on the form and position of the aerial, together with the character of the ground and the wave-length.

Norton[4] has given a series for the evaluation of $F(\omega)$ and has plotted its modulus and phase angle for a wide range of values of ω, expressed in terms of its modulus and phase angle. Burrows[5] has also plotted a function similar to the modulus, and Clemmow and Mumford have tabulated the real and imaginary parts of a related function.[7]

In some types of calculation—in particular those involving ground reflection coefficients in the form given by McPetrie[6]—it is convenient to have values of the real and imaginary parts of $F(\omega)$ expressed in terms of the real and imaginary parts of ω. Tables in this form have therefore been prepared by E. J. York, using the Ace Pilot Model, at the Mathematics Division of the National Physical Laboratory in collaboration with the Radio Research Station.

Real and imaginary components of ω are denoted by $\mathscr{R}(\omega)$ and $\mathscr{I}(\omega)$, and three Tables are given as follows:

Table 1.—For values of $\mathscr{R}(\omega)$ and $+\mathscr{I}(\omega)$ between 0 and 10 at intervals of 0·5.

Table 2.—For values of $\mathscr{R}(\omega)$ and $-\mathscr{I}(\omega)$ between 0 and 2 at intervals of 0·1.

Table 3.—For values of $\mathscr{R}(\omega)$ and $-\mathscr{I}(\omega)$ between 0 and 0·5 at intervals of 0·05.

REFERENCES

(1) SOMMERFELD, A.: "On the Spreading of Waves in Wireless Telegraphy," *Annalen der Physik*, 1909, **28**, p. 665.

(2) NORTON, K. A.: "The Propagation of Radio Waves over the Surface of the Earth and in the Upper Atmosphere: Part I," *Proceedings of the Institute of Radio Engineers*, 1936, **24**, p. 1367.

(3) MCPETRIE, J. S., and STICKLAND, A. C.: "Reflection Curves and Propagation Characteristics of Radio Waves along the Earth's Surface," *Journal I.E.E.*, 1940, **87**, p. 135.

(4) NORTON, K. A.: "The Calculation of Ground Wave Field Intensity over a Finitely Conducting Spherical Earth," *Proceedings of the Institute of Radio Engineers*, 1941, **29**, p. 623.

(5) BURROWS, C. R.: "Radio Propagation over Plane Earth—Field Strength Curves," *Bell System Technical Journal*, 1937, **16**, p. 45.

(6) MCPETRIE, J. S.: "The Reflection Coefficient of the Earth's Surface for Radio Waves," *Journal I.E.E.*, 1938, **82**, p. 214.

(7) CLEMMOW, P. C., and MUMFORD, CARA M.: "A Table of $\sqrt{(\frac{1}{2}\pi)}e^{\frac{1}{2}i\pi\rho^2}\int_{\rho}^{\infty}e^{-\frac{1}{2}i\pi\lambda^2}d\lambda$ for Complex Values of ρ," *Philosophical Transactions of the Royal Society of London, A*, 1952, **245**, p. 189.

PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

ISSUED IN THREE PARTS AS FOLLOWS:

Part A. POWER ENGINEERING (*February, April, etc.*)
Part B. RADIO AND ELECTRONIC ENGINEERING (*January, March, etc.*)
Part C. INSTITUTION MONOGRAPHS (*March and September only*)

Part B. RADIO AND ELECTRONIC ENGINEERING (INCLUDING COMMUNICATION ENGINEERING)

MAY 1955

CONTENTS OF THIS ISSUE

Bibliographical References.—It is requested that bibliographical reference to an Institution paper should always include the serial number of the paper and the month and year of publication, which will be found at the top right-hand corner of the first page of the paper. This information should precede the reference to the Volume and Part.
Example.—Smith, J.: "Reflections from the Ionosphere," *Proceedings I.E.E.*, Paper No. 3001 R, December, 1954 (102 B, p. 1234).

Published by The Institution, Savoy Place, London, W.C.2. *Telephone:* Temple Bar 7676. *Telegrams:* "Voltampere, Phone, London."
Printed by Unwin Brothers Limited, Woking and London.